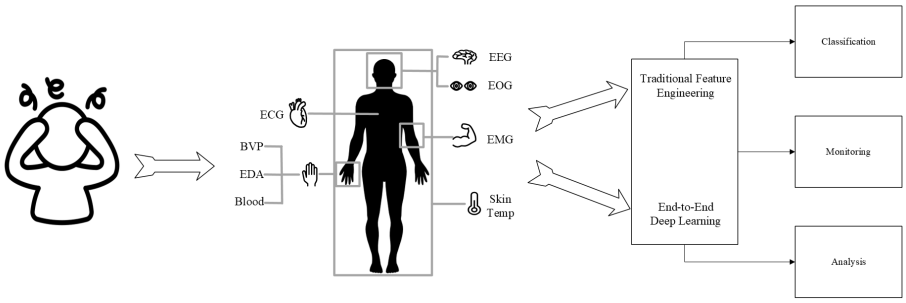
Anxiety detection using neural and physiological signals and artificial intelligence: A comprehensive review



Graphical Abstract

Anxiety Detection Using Neural and Physiological Signals and Artificial Intelligence: A Comprehensive Review

Hamidreza Ghonchi[a], Tom Foulsham[b], Saideh Ferdowsi[a]

Highlights

Anxiety Detection Using Neural and Physiological Signals and Artificial Intelligence: A Comprehensive Review

Hamidreza Ghonchi[a], Tom Foulsham[b], Saideh Ferdowsi[a]

- AI models show high accuracy in detecting anxiety from physiological signals.
- Hybrid CNN-LSTM and Transformer models represent the state-of-the-art.
- Multimodal fusion of EEG and peripheral signals yields the most robust results.
- Generalizability is limited by a reliance on small, lab-based datasets.
- Future work requires a focus on XAI, clinical validation, and ethical AI.

Anxiety Detection Using Neural and Physiological Signals and Artificial Intelligence: A Comprehensive Review

Hamidreza Ghonchi[a], Tom Foulsham[b], Saideh Ferdowsi[a]

[a] *School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester, UK*

[b] *Department of Psychology, University of Essex, Colchester, UK*

Abstract

Anxiety disorders are a significant challenge to global health, yet current diagnostic methods rely mainly on subjective and episodic assessment. The application of artificial intelligence (AI) in combination with neural and physiological signals is a promising pathway toward objective and continuous monitoring. This review provides a comprehensive evaluation of machine learning (ML) and deep learning (DL) techniques applied to electroencephalography (EEG), electrocardiography (ECG), Photoplethysmography (PPG), and electrodermal activity (EDA) in the detection of anxiety. A systematic review of the literature covering the period $2015 - 2025$ indicates that the field has evolved from traditional ML systems based on hand-crafted features to modern end-to-end deep learning schemes. Our review indicates that while classical models remain effective, hybrid models such as CNN-LSTM and more sophisticated architectures, like Transformers, consistently deliver state-of-the-art results, particularly in multimodal data integration scenarios. Despite the impressive accuracies reported, a critical examination identifies several key challenges including reliance on narrow, laboratory-trained datasets, a lack of standardized validation procedures, and limited transparency in complex models, collectively impeding clinical translation. Progress in this field requires the creation of large-scale, clinically-validated datasets; the development of fault-tolerant, generalizable, and interpretable schemes; and a transition from basic state classification to-

wards longitudinal, personalised monitoring in support of just-in-time intervention. As a comprehensive and up-to-date narrative synthesis of AI-driven anxiety detection using neural and physiological signals, this review consolidates a fragmented literature, articulates the translational gaps limiting real-world deployment, and establishes a clear research agenda for the development of clinically impactful digital mental health technologies.

Keywords: Anxiety, Machine Learning, Deep Learning, Physiological Signals, Neural Signals, EEG, HRV

---

1. Introduction

Anxiety, as a basic emotional reaction in humans, plays a vital role in helping individuals to respond to potential dangers. Although anxiety is a useful adaptation response, it crosses a clinical threshold when it is excessive, persistent, and causes interference in everyday functioning, often manifesting as a diagnosed anxiety disorder [1]. Anxiety disorders represent a substantial global public health problem, affecting an estimated 301 million individuals worldwide, representing around 4% of the total global population [2]. Anxiety disorders are among the most frequently observed psychiatric disorders, causing significant strain on individuals, healthcare infrastructure, and economies [3, 4]. This situation thus underscores the urgent need for more effective approaches to assessment and management.

The current standard way of diagnosing anxiety relies on clinical interviews, guided by criteria from the Diagnostic and Statistical Manual of Mental Disorders and subjective self-report questionnaires like the GAD-7 [5, 6]. However, these methods have inherent limitations. Not only are they essentially subjective, but they also depend on a patient's self-awareness and willingness to disclose symptoms. Additionally, these methods are vulnerable to recall bias, as patients are often required to summarize their feelings over weeks or months. Furthermore, assessments are necessarily episodic, and one short clinical visit may not reflect the dynamic, fluctuating nature of anxiety in daily life. The

stigma surrounding mental health can also deter individuals from seeking help or reporting symptoms accurately, leading to underdiagnosis and delayed treatment [7, 8].

In order to overcome these limitations, researchers are increasingly turning to artificial intelligence to develop objective, continuous, and non-invasive methods for anxiety detection. Anxiety manifests not only emotionally and cognitively but also through distinct physiological changes mediated by the autonomic nervous system. Hence, signals such as electroencephalography (EEG), electrocardiography (ECG), electrodermal activity (EDA), etc., offer objective and quantifiable biomarkers of an individual's anxious state [9]. The rapid development of wearable sensors has made it feasible to collect such data continuously in real-world settings. Machine learning (ML) and deep learning (DL) models can analyze these complex data to identify patterns related to anxiety, which could then enable early detection, personalized monitoring, and timely intervention [7, 10]. In the present study, we survey these developments and evaluate the state-of-the-art for detecting anxiety from physiological signals.

1.1. Related work and Unique Contributions

While the application of AI to mental health has been reviewed in recent years, existing surveys often differ significantly from the current work in scope or clinical focus. For instance, Kyrou et al. [11] provided an exhaustive survey of deep learning techniques for stress detection and concluded that while deep learning outperforms traditional methods, the field suffers from a lack of explainability. Similarly, Tayarani-N et al. [12] reviewed machine learning for anxiety assessment with a broad scope, including text mining. While their findings regarding social media data analysis were promising, they also identified significant limitations, particularly demographic bias and privacy concerns. Alaaeldin et al. [13] conducted a systematic analysis focused on risk-of-bias, concluding that the majority of existing studies lack rigorous external validation.

Building on this previous work, our review provides a distinct and critical

synthesis focused on bridging the gap between algorithmic performance and real-world applicability. Unlike prior reviews, this paper uniquely contributes to the literature by: (1) focusing on the latest state-of-the-art analytical pipelines proposed for analysing central (e.g., EEG) and autonomic (e.g., ECG, EDA) signals, as well as their integration within a multimodal framework for anxiety assessment; (2) explicitly conceptualising the "Anxiety and Stress" to justify the use of stress-induction protocols as valid proxies for anxiety detection while delineating their limitations; and (3) moving beyond conventional performance metrics to critically evaluate the generalizability and "in-the-wild" viability of current models.

1.2. Methodology: Search Strategy and Selection Criteria

To ensure the reproducibility of this review, we conducted a structured literature search across PubMed, IEEE Xplore, and Scopus, covering the period from January 2015 to October 2025. The search strategy employed Boolean logic to intersect the clinical domain, data modality, and computational approach. The primary search string was: (("anxiety" OR "stress") AND ("detection" OR "recognition" OR "classification") AND ("EEG" OR "ECG" OR "EDA" OR "GSR" OR "physiological signals") AND ("machine learning" OR "deep learning" OR "artificial intelligence")). Articles were selected based on specific inclusion criteria: *i)* publication in peer-reviewed journals or top-tier conference proceedings; *ii)* use of human physiological signals; and *iii)* reporting of quantitative performance metrics (e.g., accuracy, AUC). We applied exclusion criteria to filter out: *i)* studies focusing solely on text, audio, or video analysis without physiological correlates; *ii)* non-English publications; and *iii)* abstract-only papers. This process ensured a focus on high-quality, technically rigorous studies relevant to the transition from signal processing to clinical application. It is important to distinguish the scope of citations in this work: the studies detailed in Tables 3 and 4 represent the core corpus of literature selected through this systematic search for comparative evaluation. Other citations appearing in the introduction or discussions serve to provide theoretical background, clarify clinical constructs,

or support technical explanations.

The remainder of this paper is structured as follows: Section 2 details the primary data modalities and physiological justifications, along with a summary of public datasets. Section 3 provides a critical review of the core computational methods, ranging from traditional machine learning pipelines to state-of-the-art deep learning architectures. Section 4 presents a comparative performance analysis of these methods, highlighting trends in validation and accuracy. Section 5 discusses the principal findings, major challenges, and implications. Finally, the paper is concluded in section 6.

2. Definitions and Methods in Anxiety Research

2.1. Anxiety, Stress, and the State-Trait Distinction

To accurately frame the scope of this review, it is necessary to first distinguish between the psychological constructs of anxiety. Anxiety is generally categorized into state anxiety, a transient emotional reaction to a specific situation characterized by tension and autonomic arousal, and trait anxiety, a stable personality characteristic reflecting a predisposition to perceive situations as threatening. While the ultimate clinical goal is often the management of trait-based disorders, AI-based detection systems primarily operate by identifying the physiological correlates of state anxiety through moment-to-moment fluctuations in arousal [12].

This distinction elucidates the conceptual relationship between "anxiety" and "stress". Although psychologically distinct, anxiety and stress share a substantial physiological overlap. Physiologically, state anxiety manifests as an acute stress response. Whether the stressor is an immediate external pressure (defined as stress) or the anticipation of a future threat (defined as anxiety), both conditions trigger the activation of the sympathetic nervous system and the Hypothalamic-Pituitary-Adrenal axis. This results in shared fight-or-flight biomarkers evidenced by increased heart rate, reduced heart rate variability, increased skin conductance, and beta-band EEG dominance [14].

Consequently, in the machine learning literature reviewed here, studies tar-

geting "stress detection" are deliberately included alongside those targeting "anxiety". This inclusion is justified by the practical reality that inducing genuine, pathological anxiety in a controlled laboratory setting raises significant ethical and practical difficulties. Therefore, to accurately measure physiological symptoms, researchers commonly use acute stress induction protocols (e.g., the Trier Social Stress Test or mental arithmetic tasks used in datasets such as WESAD) as reliable proxies for the physiological arousal associated with anxiety [15, 11]. Throughout this review, we analyze high-quality studies targeting both constructs, treating acute stress responses as the observable, ground-truth proxies for state anxiety.

2.2. Physiological and Behavioral Modalities

While anxiety manifests through a broad spectrum of biological markers, ranging from neuroendocrine levels (e.g., cortisol) to functional brain imaging (e.g., fMRI), this review specifically prioritizes EEG, ECG, and EDA. These modalities were selected based on two critical criteria: physiological representativeness and ambulatory feasibility.

Physiological representativeness means that these signals offer a comprehensive view of anxiety: EEG reflects central nervous system activity and cognitive processing, while ECG and EDA provide robust, rapid, and reliable measurement of autonomic nervous system activation [12, 15]. Unlike fMRI or biochemical assays, which require restrictive laboratory settings, these three signals can be reliably captured using non-invasive, low-cost wearable sensors. This meets our second criterion of ambulatory feasibility. These signals are primary candidates for the development of continuous, in-the-wild monitoring systems, which represent the current trajectory of the field toward accessible digital health solutions [7].

This review focuses on the use of AI to detect anxiety through objective signals from the brain and body. In the following subsections, we briefly introduce these key signals.

2.2.1. Electroencephalography (EEG)

Electroencephalography is a neurophysiological technique that measures the brain cells' electrical activity using electrodes placed on the scalp [9]. It provides a non-invasive gauge of arousal states based on neural dynamics. Anxiety is consistently associated with specific changes in neural oscillations, particularly a decrease in alpha band ($8 - 12$ Hz) power, which is dominant during relaxed states, and an increase in beta band ($12 - 30$ Hz) power, which is linked to alertness and active thinking [16, 17]. Researchers analyze features such as Power Spectral Density (PSD) within these bands, functional connectivity between brain regions, and hemispheric asymmetry to identify the neural signatures of anxiety [18, 19].

2.2.2. Electrocardiography (ECG)

Electrocardiography measures the electrical activity of the heart [9]. The heart's rhythm is regulated by the autonomic nervous system and is highly sensitive to emotional arousal, making ECG a powerful tool for detecting anxiety. Anxious states typically trigger the sympathetic "fight-or-flight" response that leads to an increased heart rate and, crucially, a reduction in Heart Rate Variability (HRV) [20, 21]. HRV, which is the variation in time between consecutive heartbeats, is the most common and robust metric derived from ECG for anxiety assessment, with lower HRV indicating reduced psychological and physiological adaptability [22].

2.2.3. Photoplethysmography (PPG)

Photoplethysmography is an optical sensing technique widely used in smartwatches and fitness trackers to quantify blood volume variations within the peripheral microvasculature [23]. By emitting light into the skin, PPG enables the estimation of heart rate and heart rate variability, often referred to as Pulse Rate Variability (PRV) within this context. It thus offers a practical and non-invasive alternative to ECG in ambulatory monitoring. Similar to HRV derived from ECG, reduction in PRV correlates with elevated stress and anxiety levels,

indicating increased sympathetic nervous system activity[24, 25].

2.2.4. Electrodermal Activity (EDA) or Galvanic Skin Response (GSR)

EDA, also referred to as GSR, measures variations in the skin's electrical conductivity [18]. This phenomenon reflects the activity of the sweat glands, which are governed solely by the sympathetic nervous system. This direct association renders EDA an exceptionally sensitive and specific indicator of both emotional and physiological arousal [20]. States of anxiety result in heightened sweat secretion, which correlates with significant increases in the baseline tonic skin conductance level (SCL) as well as the frequency and amplitude of phasic skin conductance responses (SCRs) [24].

2.3. Publicly Available Datasets

The progress of research in anxiety detection using machine learning and deep learning has been significantly facilitated by the availability of publicly accessible datasets. There are several publicly-available benchmark datasets in the field. These datasets comprise multimodal physiological signals (such as EEG, ECG, and EDA), recorded while participants are exposed to stimuli designed to elicit specific emotional states, such as anxiety, fear, or stress. Common protocols range from watching emotionally charged video clips (e.g., in DEAP and SEED) to undergoing face-to-face psychological stimulation (e.g., in DASPS) or performing stressful cognitive tasks (e.g., Sam40). The most useful datasets for training and comparing supervised machine learning models provide these measures alongside labels of the emotional state or anxiety level, as derived from self-reported questionnaires. A summary of the most relevant publicly available datasets for anxiety and emotion recognition research is provided in Table 1.

Although these datasets are invaluable for benchmarking algorithms, it is critical to acknowledge their common limitations. Some datasets, such as DEAP and WESAD, were collected in controlled laboratory environments using standardized emotional stimuli on healthy, often young, student participants. Therefore, machine learning models trained on this data may not generalize well to

Table 1: A summary of key publicly available datasets for anxiety, stress, and emotion research

Dataset Name	Primary Focus	Task / Stimulus	Data Modalities Recorded	Number of Subjects	Ground Truth Labeling Method
Physiological Datasets (EEG, ECG, etc.)					
DASPS (A Database for Anxious States) [26]	Anxiety States	Imagining anxious scenarios guided by a psychotherapist	EEG, ECG	23	Self-Assessment Manikin (SAM), HAM-A scores
WESAD (Wearable Stress and Affect Detection) [27]	Stress & Affect	Mental arithmetic, public speaking, watching amusing video	ECG, EDA, EMG, RESP, TEMP, BVP	15	Self-report (stress level), condition-based labels
DEAP (Database for Emotion Analysis using Physiological signals) [28]	Emotion Recognition	Watching 1-minute music video excerpts	EEG, EOG, EMG, EDA, RESP, TEMP	32	Self-assessment (Valence, Arousal, Dominance)
SEED (SJTU Emotion EEG Dataset) [29]	Emotion Recognition	Watching emotional movie clips (positive, neutral, negative)	EEG	15	Emotion category labels (happy, sad, fear, etc.)
AnxiECG-PPG [30]	Anxiety States	Watching emotion-inducing videos, physical stress (hand immersion)	ECG, PPG	47	Self-assessment (Valence, Arousal, Anxiety)
VerBIO [31]	Public Speaking Anxiety	Giving speeches in real-life and Virtual Reality (VR)	Audio, ECG, EDA	55	Self-report (Trait/State Anxiety, Stress Ratings)
BESST (Brno Extended Stress and Speech Test) [32]	Cognitive & Physical Stress	Cognitive tasks (Reading Span), physical stress (Hand Immersion)	Speech, ECG, EDA, Skin Temp, Video	90	Questionnaire, task-based labels
PROPER (Personality Recognition on Public Speaking) [33]	Public Speaking	Giving a speech to a real audience	EEG, Wearable sensor data	40	Big Five personality test (for correlation)
SAM 40 [34]	Induced Stress	Stroop test, arithmetic task, mirror image recognition	EEG	40	Verbal feedback, task-based labels
EEG During Mental Arithmetic Tasks [35, 36]	Stress	Mental arithmetic tasks	EEG	66	Task-based labels (rest vs. task)
HR-EEG4EMO [37]	Emotion Recognition	Watching neutral and emotional film excerpts	High-resolution EEG	40	Stimulus-based labels (neutral vs. emotional)
MODMA (Multi-modal Open Dataset for Mental-disorder Analysis) [38]	Mental Disorder (Depression)	Viewing Fear, Sad, Happy, Neutral stimuli	EEG, Speech	53	Diagnosis-based (Depression vs. Control)
Preschool Children Anxiety Dataset [39]	Anxiety Disorder Diagnosis	Clinical interviews, parent-report questionnaires	Behavioral, psychophysiological data	193	Clinical diagnosis (GAD, SAD, Panic Disorder)
Anxiety Phase Dataset [40]	Phobic & Social Anxiety	Exposure to anxiety-provoking stimuli (e.g., bug-related)	ECG, EDA (Torso/Wrist/Leg)	95	Questionnaire, condition-based labels
Text-Based Datasets					
Dreaddit: A Reddit Dataset for Stress Analysis [41]	Stress (from text)	N/A (analysis of existing online posts)	Text (Reddit posts)	N/A	Subreddit labels (e.g., posts from r/anxiety vs. r/happy)
Stress Detection from Social Media Articles [42]	Stress (from text)	N/A (analysis of existing online posts)	Text (Reddit, Twitter posts)	N/A	Keyword and content-based labels

diverse and real-world clinical populations experiencing spontaneous and idiopathic anxiety. Furthermore, the relatively small number of subjects in many datasets (often fewer than 50) poses a significant challenge for training complex deep learning models and increases the risk of overfitting which restricts the robustness of these models.

Recognizing these gaps, the field is slowly moving towards creating more valid datasets. For instance, datasets like VerBIO [31] and the data collection methodology described by Ihmig et al. [43] capture physiological responses during more realistic tasks such as public speaking or virtual reality exposure therapy. The development of datasets that include clinically diagnosed patient populations, longitudinal data collected over days or weeks, and data captured in-the-wild using consumer-grade wearables remains a critical and urgent priority for the research community. Such datasets will be essential for validating the clinical utility of AI models and moving them from the laboratory to real-world applications.

2.4. Psychological Questionnaires as Ground Truth

An essential step in developing supervised machine learning models for anxiety detection is the creation of an accurately labeled dataset. In this context, researchers mostly use validated psychological questionnaires and rating scales in order to establish the ground truth. These self-report or clinician-administered instruments are used to assess a participant's subjective emotional state, with the resulting scores serving as the labels (e.g., anxious vs. non-anxious, or a numerical severity score) for the corresponding physiological data segments. The most widely used questionnaire is the State-Trait Anxiety Inventory (STAI), which distinguishes transient situational anxiety (state) and an individual's general disposition (trait), and the Generalized Anxiety Disorder-7 (GAD-7), a brief scale widely used for screening clinical symptoms [44, 6]. For studies targeting specific phobias or situations, more specialized scales such as the Personal Report of Public Speaking Anxiety (PRPSA) are also employed [45].

As an alternative to subjective, state-based labeling, a clinically-relevant

approach is to establish ground truth using a formal clinical diagnosis. This method is indeed feasible and is employed in a growing number of studies where the goal is to classify the person themselves. For example, studies by Luo et al. [46], Al-Ezzi et al. [47], and Park et al. [10] successfully trained models to distinguish between individuals with diagnosed Generalized Anxiety Disorder (GAD) or Social Anxiety Disorder (SAD) and healthy individuals. Although a clinical diagnosis provides a more stable and objective gold standard label, it also presents its own challenge. For example, a person with an anxiety disorder is not physiologically anxious every second of the day, creating a temporal ambiguity when labeling continuous time series data. Therefore, understanding which type of ground truth is used in a study, a momentary, questionnaire-based state score or a stable, diagnosis-based trait label, is crucial for critically evaluating the claims and applicability of any AI model for anxiety detection. A comprehensive list of questionnaires used for measuring (ground truth) anxiety and stress is provided in Appendix A.

3. Computational Methods for Anxiety Detection

The methodologies employed for automated anxiety detection range from traditional machine learning techniques, which rely on handcrafted features, to deep learning models capable of automatic representations learned from raw data. This section reviews these core computational paradigms, highlighting their application to physiological signals for classifying anxious states. A summary of the most commonly engineered features for anxiety detection is provided in Table 2.

3.1. The Traditional Machine Learning Pipeline

Traditional machine learning represents the foundational approach to anxiety detection. This paradigm typically involves a two-stage process: first, extracting a set of predefined features from the raw input data (e.g., physiological signals) based on domain knowledge; second, training a standard classification algorithm on these extracted features.

Table 2: Commonly used features for traditional machine learning-based anxiety detection

Biological Signal	**Feature Category**	**Example Features**
EEG	Time-Domain	Mean, Standard Deviation, Zero-Crossing Rate, Hjorth Activity, Mobility, Complexity
	Frequency-Domain	PSD (Delta, Theta, Alpha, Beta, Gamma), Spectral Centroids
	Time-Frequency Domain	DWT Coefficients, Shannon Entropy of DWT
	Emotional Model	Valence, Arousal Scores
ECG	HRV - Time-Domain	RMSSD, SDNN, pNN50, MeanNN
	HRV - Frequency-Domain	LF Power, HF Power, LF/HF Ratio
	HRV - Non-Linear	Sample Entropy, SD1, SD2
	ECG Peaks Interval	Mean RR Interval, SD of RR Intervals
PPG	HRV - Time-Domain	RMSSD, SDNN, pNN50
	HRV - Frequency-Domain	LF Power, HF Power, LF/HF Ratio
	Pulse Wave Characteristics	Systolic Peak Amplitude, Diastolic Peak Amplitude, Pulse Width
EDA	Statistical	Mean Peak Width, Standard Deviation of Peak Amplitude
	Statistical	Mean EDA, Standard Deviation of EDA, Peak-to-Peak Amplitude
	SCL Features	Mean SCL, Maximum SCL
	SCR Features	Mean SCR Amplitude, Number of SCRs, Rise Time
	Time-Domain	Sum Rise Time, Sum Fall Time

3.1.1. Feature Extraction Engineering and Selection

The success of traditional machine learning models in anxiety detection is heavily dependent on the quality of hand-crafted features extracted from physiological signals. Through feature engineering, raw time series are converted into structured features that provide discriminative information for anxiety classification. The feature sets are typically tailored to the specific characteristics of each data modality.

For EEG, the features illustrating anxiety are most often extracted from the frequency and time-frequency domain. The most popular feature is Power Spectral Density (PSD), which is calculated for typical neural oscillation bands, including delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma ($>$ 30 Hz), as the relative power within bands is highly correlated with shifting states of relaxation and arousal [48, 19]. Besides spectral power, other common features include hemispheric asymmetry measurements in the form of the Rational Asymmetry (RASM) and the Frontal Asymmetry Index (FAI), examining the levels of activity in the left and right cerebral hemispheres. For the purpose of signal complexity evaluation, various types of entropy, such as Fuzzy Entropy, are used as well [18]. Since EEG is a non-stationary signal, time-frequency features estimated using the Discrete Wavelet Transform (DWT) are often used to derive statistical moments from the decomposition coefficients, enabling the detection of transient changes that may be overlooked under standard spectral analysis [49].

For cardiac signals acquired via ECG and PPG, analysis generally centers on HRV and PRV, respectively. Commonly extracted features aim to quantify the autonomic nervous system's influence on heart rate dynamics. These are described below. Time-domain features are statistical measures of the intervals between heartbeats (normal-to-normal (NN) intervals). Key examples include the standard deviation of NN intervals (SDNN), reflecting overall HRV; the root mean square of successive differences (RMSSD), capturing short-term and high-frequency variations; and pNN50, the percentage of adjacent NN intervals that

differ by more than $50ms$ [50, 51].

Frequency-domain features quantify the distribution of HRV power across spectral bands. The low-frequency band (LF:0.04–0.15 Hz) is associated with both sympathetic and parasympathetic activity, while the high-frequency band (HF: 0.15–$0.4Hz$) reflects parasympathetic (vagal) activity. The LF/HF ratio is a widely used yet often debated marker of sympathovagal balance, with higher values typically indicating increased sympathetic dominance, which is commonly observed in anxious states [52].

Non-linear features capture the complexity and regularity of the heart rate pattern, providing insights beyond linear statistics. Common examples include Sample Entropy (SampEn) and Approximate Entropy (ApEn), which measure predictability, and geometric descriptors from Poincaré plots, such as SD1 (reflecting short-term variability) and SD2 (reflecting long-term variability) [53, 54, 50].

For EDA, key features include statistical properties of the raw signal (e.g., mean and standard deviation) and, more specifically, metrics derived from its two principal components. These metrics are designed to characterize both the slowly varying tonic component and the rapidly fluctuating phasic responses of the signal, which reflect sympathetic arousal. Tonic features are derived from the SCL, the baseline level of conductivity. The mean SCL over a window is a common feature indicating overall arousal. Phasic features are derived from SCRs, which manifest as brief, transient increases in conductivity. Key SCR features include their frequency of occurrence, mean amplitude, total count within a given time window, and rise time (the time from onset to peak) [55].

Following this extensive extraction process, a feature selection step is often crucial to reduce dimensionality, remove redundant information, and prevent model overfitting. Techniques such as Recursive Feature Elimination (RFE), statistical tests (e.g., ANOVA), or wrapper-based methods are commonly applied to identify the most salient subset of features for the final classification task [19, 56].

3.1.2. Traditional Machine Learning Classifiers

After extracting a set of discriminative features, different supervised machine learning classifiers are trained to differentiate among anxious states. Although many algorithms have been evaluated, some have established themselves as particularly powerful for anxiety detection.

Support Vector Machines (SVMs) are frequently used due to their effectiveness in high-dimensional feature spaces. In this framework, the kernel function serves as a key hyperparameter and plays a critical role in determining model performance. The most common kernels used by SVM are Linear (LBF) and Radial Basis Function (RBF) kernels. For instance, Sinche et al. [57] demonstrated that an SVM classifier achieved very high accuracy (99.84%) in binary stress detection using ECG-derived features from consumer-grade heart rate sensors. SVMs have demonstrated effectiveness in multi-class classification problems as well. For example, Kang et al. [58] used a one-against-all SVM to classify four stress-related emotional states based on ECG features, and demonstrated that an ensemble SVM-Naive Bayes outperformed other models. Szakonyi et al. [59] further explored SVMs alongside other approaches for stress detection based on HRV, while Andrikopoulos et al. [60] reported that an SVM-based model yielded the highest performance (81.6% accuracy) for detecting adult ADHD, a condition often associated with anxiety, using multimodal wearable data. Although SVMs are powerful, these methods are increasingly used as a baseline for comparison against more complex ensemble methods [51].

Ensemble methods, which combine the outputs of multiple base estimators, consistently rank among the top-performing traditional models. Their ability to reduce variance and handle complex, non-linear interactions makes them well-suited for noisy physiological data. Tree-based ensemble methods have become particularly prevalent, with Random Forest (RF) standing out as a widely preferred classifier in this area, mainly because of its reliability and its ability to rank features by importance. According to research performed by Muhammad et al. [19], RF consistently surpasses SVM, k-Nearest Neighbors (kNN), and

Multilayer Perceptron (MLP) methods in tasks involving the classification of four anxiety levels based on EEG data. Similarly, Vaz et al. [56] achieved an outstanding ROC-AUC of 0.980 using an RF model coupled with data balancing techniques. More sophisticated ensemble methods, such as eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost, are increasingly gaining recognition for their strong predictive performance. Dalmeida et al. [51] and Park et al. [10] have identified these models as leading options for predicting symptoms related to stress and social anxiety, respectively. The effectiveness of these advanced ensembles was further demonstrated by Luo et al. [46], who applied CatBoost and LightGBM to EEG-derived functional connectivity features to classify Generalized Anxiety Disorder into four severity levels. CatBoost delivered accuracy of between 95.5% and 98.5%. Moreover, even a basic Decision Tree, as implemented by Tsai et al. [50], and Zeng et al. [61], can be remarkably effective, particularly when developing interpretable models for assessing stress and detecting sex-specific fatigue.

Boosting algorithms, such as AdaBoost, are another powerful class of ensemble techniques. Vaz et al. [56] highlighted that an AdaBoost classifier achieved the highest F1-score and accuracy on their anxiety classification task when combined with SMOTE for data balancing. Aldayel et al. [48] also included AdaBoost Bagging, and Gradient Bagging in their comprehensive comparison of classifiers for EEG-based anxiety detection.

Other classifiers have also been explored to provide a more comprehensive comparative picture. Al-Nafjan et al. [62] employed kNN to achieve high accuracy (96.9%) using statistical features achieved from GSR data, outperforming both SVM and RF in that specific application. However, kNN is also used as a baseline by many other researchers, including Szakonyi et al. [59] and Aldayel et al. [48]. Similarly, Linear Discriminant Analysis (LDA) is frequently included for benchmarking, as seen in the work of Bilucaglia et al. [63] and Vaz et al. [56]. Finally, simpler models like Logistic Regression and Naive Bayes are often included in initial comparative analyses but are generally superseded by more

complex models in final results [58, 64].

Overall, although a variety of classifiers are viable, the research trend in traditional machine learning for anxiety detection clearly indicates that tree-based ensemble methods, particularly Random Forest and Gradient Boosting variants, offer the most robust and accurate solutions across different data modalities and datasets.

3.2. Deep Learning Approaches

Researchers have increasingly shifted towards using deep learning models to overcome the limitations of traditional classifiers and capture more complex, non-linear patterns. While deep learning is frequently associated with end-to-end learning, where models extract hierarchical representations directly from raw data, the literature reveals a more nuanced spectrum of applications. To avoid conceptual confusion, deep learning methods in anxiety detection can be categorized into three distinct operational frameworks: (1) Feature-based deep learning, where traditional hand-crafted features (e.g., spectral power or effective connectivity) are fed into deep neural networks to leverage their advanced classification capabilities [65, 66]; (2) Representation-based hybrid learning, which transforms 1D signals into engineered representations (e.g., 2D spectrograms, scalp maps, or autoencoder latent spaces) prior to deep spatial/temporal processing; and (3) Strict end-to-end learning, which operates directly on raw or minimally processed time-series data to discover novel, entirely data-driven biomarkers. Different deep learning architectures are chosen based on this input data formulation.

3.2.1. Deep Learning Architectures for Sequential Data: RNN, LSTM, and GRU models

Recurrent Neural Networks (RNNs) and their more advanced variants, Long Short-Term Memory (LSTM) as well as Gated Recurrent Units (GRU), are naturally suited for modelling physiological signals, which are fundamentally time series. The input to recurrent neural networks can vary depending on the approach. In some hybrid models, the input is a sequence of hand-crafted HRV

features [67, 68]. In many of the end-to-end systems, meanwhile, raw or minimally processed 1D time series, such as the sequence of ECG voltage values or RR intervals, are fed directly into the network [69]. This causes the model to implicitly learn temporal patterns without relying on predefined feature extraction. These models have internal memory mechanisms which allow them to capture temporal dependencies and understand how a signal evolves. This makes them ideal for tasks like anxiety and stress detection.

Combining the strengths of traditional signal processing with temporal deep learning is a common application to process sequences of hand-crafted features. For example, Zhong et al. [67] used a deep GRU network to classify four levels of psychological stress by feeding it a sequence of seven HRV features extracted over time. Similarly, Liu et al. [68] employed a Bi-directional LSTM (BiLSTM) to process Differential Entropy (DE) features from EEG signals as part of a depression prediction model. They demonstrated the architecture's capacity to learn from sequences of pre-extracted neural features.

Together, these findings indicate that such architectures can be extended to minimally processed data, enabling the network to learn temporal patterns directly from the input. LSTMs are widely used for this purpose. Masuda et al. [70] utilized an LSTM layer to process features extracted by a CNN from raw multi-channel neural and physiological data (i.e. EEG and EMG), demonstrating effective classification of four distinct fear levels. BiLSTM variants, which process sequences in both forward and backward directions to capture richer contextual information, have also shown strong potential. Roy et al. [49] incorporated a BiLSTM layer into a hybrid model to learn from decomposed EEG signals, while Ahmed et al. [65] reported that a BiLSTM-kNN model achieved outstanding accuracy (97.88%) for detecting Obsessive-Compulsive Disorder (OCD) from EEG-derived features. The superiority of this architecture was further demonstrated by Joshi et al. [71]. After extracting novel Higher-Order Statistics (HOS) features from EEG signals in the DASPS dataset, they found that a BiLSTM network significantly outperformed both SVM and kNN classifiers trained on the same feature set. This highlights the power of recurrent

architectures in classifying a range of psychiatric conditions from neurophysiological time series data.

3.2.2. Deep Learning Architectures for Spatiotemporal Data: CNNs and Hybrids

Convolutional Neural Networks (CNNs), the cornerstone of deep learning for image analysis, have been adapted with great success to physiological signal processing. Instead of processing raw 1D time series directly, the signals can be transformed into 2D representations that serve as the input features. This preprocessing step is crucial and defines the type of features the CNN will learn. Common input representations include spectrograms, which are time-frequency images created using techniques like the Short-Time Fourier Transform (STFT) or Wavelet Transforms, visualizing how the frequency content of a signal (e.g., EEG or ECG) changes over time [72]. Another common representation involves topographical maps, which are 2D images that represent the spatial distribution of EEG activity across the scalp. This preprocessing allows the CNN to learn spatial patterns between electrodes. This method, as employed by Ghonchi et al. [73], converts 1D channel data into a sequence of 2D "scalp map" images that preserve the spatial topology of the electrodes and allow a CNN to directly learn from the spatial relationships of brain activity. Finally, Poincaré plots, which are visual representations of HRV where RR intervals in ECG signals are plotted against subsequent intervals, can be converted into images and used as input to a CNN to learn non-linear dynamics [74, 75].

Standalone CNNs can be highly effective. Dar et al. [76] used a 2D-CNN on EEG data arranged in a 2D grid for emotion classification. In a novel approach, Baygin et al. [77] used a CNN to classify images generated using a hand-crafted Probabilistic Binary Pattern (PBP) from ECG signals. They achieved over 98.5% accuracy in 4-class anxiety detection. This strategy of converting 1D signals into 2D images is gaining significant traction; Ebrahimpour et al. [78] also successfully applied this by transforming 5-second ECG segments into time-frequency images via a Continuous Wavelet Transform (CWT). By feeding these images into a powerful ResNet50 CNN, they achieved 94.92% ac-

curacy in classifying Post-Traumatic Stress Disorder (PTSD). This indicates that rich diagnostic information exists in the ECG waveform's morphology beyond standard HRV. Another innovative approach is seen in the MVCA-Net proposed by Ghonchi et al. [79], which uses parallel convolutional pathways with varying kernel sizes to automatically extract multi-scale frequency features directly from EEG signals. They used attention mechanisms to enhance performance. However, the most powerful and prevalent architecture in this domain is the hybrid CNN-LSTM model which integrates the complementary strengths of both components. In a hybrid model, the CNN serves as an automated feature extractor that captures salient spatial or spectral representations, while the subsequent LSTM (or GRU) layer models the temporal dependencies among these learned features. This end-to-end spatiotemporal approach has consistently delivered state-of-the-art results in laboratory settings. Nevertheless, these high-performance metrics must be interpreted with a critical lens; many of these studies rely on small participant pools and extensive data augmentation, which can artificially inflate accuracy and introduce a high risk of overfitting.

Several studies have highlighted the effectiveness of hybrid architectures for stress, fear, and anxiety detection across diverse physiological modalities. In the context of stress and fear detection, Ghonchi et al. [73] is a prime example. We proposed a hybrid CNN-RNN framework that processed sequences of the generated scalp maps, achieving accuracies of 94.24% for binary and 92.58% for four-class anxiety detection on the DASPS dataset. Similarly, Tanwar et al. [80] developed a CNN-LSTM to recognize stress levels using multimodal wearable sensor data, including ECG, EMG, and EDA signals. Masuda et al. [70] further extended the paradigm by introducing a multi-input CNN-LSTM, where CNNs first extracted features from raw multichannel physiological data before an LSTM layer classified four levels of fear with high accuracy.

For High-Accuracy ECG Analysis, Kang et al. [72] applied a modified CNN-LSTM ensemble to ECG spectrograms. They achieved a stress classification accuracy of 98.3%, demonstrating a significant improvement over previous meth-

ods.

To fuse signals from multiple modalities, Xiang et al. [72] designed a multimodal model (MMFD-SD) using parallel CNN branches to process features from accelerometer, EDA, HR, and temperature data. They showed robust performance in occupational stress detection. Kim et al. [81] further developed this by using a multi-column CNN that fused features from both EEG and GSR spectrograms. They achieved an outstanding AUROC of 0.954 for stress classification in a realistic VR interview paradigm.

When applied to psychiatric condition diagnosis, the hybrid model's utility extends beyond stress. Ahmed et al. [65] found that a CNN-LSTM outperformed other models for diagnosing adjustment disorder from EEG signals. Similarly, Al-Ezzi et al. [82] utilized a hybrid CNN-LSTM to classify the severity of Social Anxiety Disorder into four distinct levels. By training the model on features derived from the brain's effective connectivity networks, they achieved a high accuracy of 92.86% and demonstrated that the default mode network (DMN) was a key predictive region. Roy et al. [49] created a hybrid model integrating a 1D-CNN with BiLSTM and GRU layers (CBGG) to detect psychological stress from decomposed EEG signals. They reported a classification accuracy of 98.10%.

These studies collectively show that by converting time series problems into a spatiotemporal format, hybrid models, particularly the CNN-LSTM, can effectively learn the complex, intertwined patterns indicative of anxiety and its related states.

3.2.3. Advanced and Unsupervised Architectures: Transformers, GNNs, Autoencoders

Beyond the well-established CNN- and RNN-based architectures, some researchers have shifted toward more advanced models designed to tackle specific challenges like long-range temporal dependencies, inter-channel relationships, and the need for unsupervised feature learning.

Transformers and Attention Mechanisms, originally developed for natural

language processing, are increasingly applied to physiological signals due to their ability to model long-range dependencies more effectively than RNNs. While RNNs process data sequentially and may suffer from the loss of early information in long sequences, transformer models utilize a self-attention mechanism to simultaneously evaluate the relative importance of all time steps. These advantages were highlighted in recent work on stress detection [83, 84]. The self-attention mechanism allows the model to identify salient physiological events (e.g., a sporadic spike in skin conductance or a specific EEG burst) regardless of where they occur in the time series, making them ideal for capturing the complex, non-stationary dynamics of anxiety [11]. Behinaein et al. [83] developed a novel Transformer-based architecture to perform end-to-end stress detection directly from raw ECG signals. The results they achieved were comparable to the state-of-the-art methods without any manual feature engineering. Arjun et al. [84] further advanced this direction by demonstrating that a Vision Transformer (ViT) applied directly to raw 1D EEG signals could outperform a ViT applied to 2D time-frequency images. This method achieved 99.4% accuracy and highlighted the power of the attention mechanism on raw temporal data. However, such exceptionally high accuracies on restricted datasets must be interpreted cautiously due to potential subject-dependent bias. Additional hybrid approaches have employed attention mechanisms. For example, Liu et al. [85] incorporated a transformer module into their ERTNet architecture for emotion recognition via EEG, while Ghonchi et al. [79] combined a transformer with a Squeeze-and-Excitation (SE) attention block within their convolutional MVCA-Net architecture for dynamically focusing on the frequency-specific elements within EEG signals for the task of anxiety detection. Building upon this architecture, Liu et al. [86] proposed the EEGMind-Transformer, an end-to-end novel model utilizing dynamic temporal graph attention along with hierarchical graph representation. This advanced configuration allows the model to skillfully extract the complex spacial and temporal dynamics within the brain from minimally preprocessed EEG input. They achieved state-of-the-art performance, reaching an accuracy of 92.5% on publicly released data sets. While estab-

lishing the immense possibility of dedicated transformer-based architectures, it must be noted that these models carry significant computational costs and require massive data volumes to generalize without overfitting, currently limiting their deployment in low-power wearable systems.

Graph Neural Networks (GNNs) represent a new frontier for EEG analysis by explicitly modeling the topological relationships among sensors. Standard CNNs typically treat EEG channels as a flat 2D grid, which ignores the 3D structure of the brain. In contrast, GNNs represent electrodes as nodes and their connections, defined by physical proximity or functional connectivity, as edges. This allows the model to learn not only from the signals at individual electrodes, but also from the synchronized activity patterns across brain regions. While the current literature lacks extensive empirical evidence proving GNN superiority over CNNs specifically for anxiety detection, this network-based approach holds strong forward-looking potential for capturing the network-level dysregulation characteristic of anxious states. For instance, in related affective computing domains, Liu et al. [68] merged a Graph Convolutional Network (GCN) with a BiLSTM in their DBGCN to integrate spatial topology with time-frequency features. Their model showed significantly improved depression prediction accuracy. Guo et al. [87] developed a complex model, termed SAG-CET, combining a GCN with a cross-EEG transformer to jointly model both inter-channel spatial relationships and multiscale temporal features. They achieved state-of-the-art performance on emotion recognition tasks. Although most commonly applied to depression or emotion analysis, Tian et al. [88] demonstrated GNN's utility for unsupervised anomaly detection in EEG data, a technique with clear theoretical potential for identifying neural dysregulation in anxiety disorders.

Autoencoders are primarily used for unsupervised feature learning. These neural networks are trained to reconstruct their input, forcing them to learn a compressed and salient representation of the data in a central bottleneck layer. This means that the input to the autoencoder is the raw signal itself (e.g., a segment of a GSR signal), and the output serves as a learned feature vector for subsequent classification [62]. This learned representation can then be used as a

powerful feature set for a simple classifier. Shikha et al. [89], in a broader review, highlighted the use of a Stacked Sparse Autoencoder (SSAE) for unsupervised feature learning from EEG, which was then fed to a Decision Tree to classify anxiety levels. In a direct application, Al-Nafjan et al. [62] used a 14-layer autoencoder to extract features from GSR signals. They found that a simple kNN classifier using these autoencoder-derived features achieved an accuracy of 98.2%, outperforming the same classifier trained on traditional statistical features which achieved 96.9% and thereby demonstrating the superiority of learned representations.

4. Comparative Analysis and Performance

Evaluating the performance of different anxiety detection methods is inherently complex, as outcomes depend heavily on the dataset, the data modalities, the validation methods employed, and the performance metrics reported. Although many studies in the literature report high classification accuracies, a direct comparison of them without considering their context can be misleading. To facilitate a more nuanced and critical analysis, Tables 3 and 4 provide a comparative summary of the primary studies identified through our systematic literature search. These tables do not provide an exhaustive list of all published studies, but rather a representative subset of high-quality research that illustrates the performance of specific architectures, from traditional machine learning pipelines to end-to-end deep learning systems, across various modalities and experimental protocols.

A critical review of the findings summarized in Tables 3, 4 and 5 reveals several important trends and considerations for the field.

4.1. The Challenge of Comparing Performance Metrics and Validation Schemes

A key observation from the tables is the considerable variance in reported performance, with accuracies up to 99%. This variation often reflects differences in experimental setup rather than the inherent superiority of any particular model.

Table 3: Comparative performance of selected anxiety and stress detection studies that used traditional machine learning

Study (Ref)	Data Modality(s)	Dataset	# Subjects	Model Architecture	Key Performance (Metric)	Clinical Context / Limitations
Traditional Machine Learning Approaches						
Muhammad et al. [19]	EEG	DASPS (Public)	23	• **Features:** PSD (Welch method), FFT. • **Selection:** Recursive Feature Elimination (Wrapper). • **Classifier:** Random Forest (RF) (Best performing). Also tested SVM, MLP, kNN, DT.	**Best Metrics:** • **2-Class (HAM-A labels):** Accuracy: 94.90%, F-value: 0.93, Kappa: 0.88. • **4-Class (HAM-A labels):** Accuracy: 92.74%, F-value: 0.90, Kappa: 0.86. • **Best Class Metrics (Severe Anxiety):** Precision: 0.96, Recall: 0.96.	• **Context:** Lab-based exposure therapy (psychological stimulation). Used HAM-A and SAM scales for ground truth. • **Limitations:** Small sample size (23); limited demographic (only 30 years old, one ethnicity); consumer-grade headset (though validated); only visual/auditory stimuli used.
Aldayel et al. [48]	EEG	DASPS (Public)	23	• **Features:** DWT (Discrete Wavelet Transform) performed best (better than PSD). • **Classifier:** Random Forest (RF) achieved the highest results. They also tested Ensemble methods (AdaBoost, Gradient Bagging) and standard ML (SVM, KNN, LDA).	**Best Metrics (RF + DWT + HAM-A labels):** • Accuracy: 87.50% • Precision: 87.65% • Recall: 87.50% • AUROC (AUC): 0.8781 • Kappa: 0.757 (Acc)	• **Context:** Exposure therapy (Flooding) using visual stimuli. Compared subjective labels (SAM) vs. clinical scale labels (HAM-A). HAM-A was superior. • **Limitations:** Class imbalance (required SMOTE), "Not yet ready for clinical use," individual variability in EEG signals.
Kang et al. [58]	ECG	CLAS (Private)	31	• **Features:** R-S peak, R-R interval, Q-T interval. • **Classifier:** Ensemble of SVM + Naive Bayes (SVM-NB).	**Best Metrics:** • Accuracy: 97.6% • Precision: 97.5% • Recall: 97.4% • AUC: 97.9%	• **Context:** 4 emotional states (Picture, Music, Stroop, Math). Classifying Stress (levels 3-4) vs. Non-Stress (levels 1-2). • **Limitations:** Focuses on short-term induced stress; dataset limited to 31 subjects.
Sinche et al. [57]	ECG, PPG	Private	20	• **Features:** Time-domain (SDNN, RMSSD, pNN50) + PAT features. • **Classifier:** SVM (SMO-RBF) performed best compared to Decision Tree (J48) and Vote meta-classifier.	**Best Metrics:** • Accuracy: 99.84% (Using All features: ECG+PPG+PAT). • Modality Comparison: PAT features alone achieved 93.53% Acc; ECG alone 84.51%; PPG alone 64.14%.	• **Context:** Binary classification: "Stressed" (before an exam) vs. "Non-stressed" (after the exam). Validated using "Cognitive Test Anxiety Scale". • **Limitations:** Small sample size (20); frequency domain analysis excluded; PPG sensor (MAX30100) found less reliable than ECG.
Vaz et al. [56]	ECG, EDA, EMG	WESAD (Public)	15	• **Preprocessing:** Filtering, Segmentation. • **Features:** RFECV (Recursive Feature Elimination). • **Data Balancing:** SMOTE, ADASYN, Borderline SMOTE 2. • **Classifier:** AdaBoost (ADB) with SMOTE achieved the highest Accuracy and F1. (Random Forest achieved the highest AUC).	**Best Metrics (for ADB + SMOTE):** • Accuracy: 92.0% • F1-score: 0.864 • Precision: 0.875 • Recall: 0.867 • AUC: 0.961	• **Context:** Unlike other studies using WESAD, this study only used the "Neutral" baseline condition and classified anxiety based on STAI-S scores (Low/Mod vs High). • **Limitations:** Small sample size (15); data imbalance (requires synthetic oversampling); "Neutral" state might be contaminated by anticipation of future stress tasks.
Dalmeida et al. [51]	ECG, PPG/Heart Rate	DRIVE-DB (Public), Apple Watch Data (Private)	27 (DRIVE-DB), 4 (Apple Watch Data)	• **Features:** HRV (Time and Frequency domain). • **Classifiers:** Gradient Boosting (GB) and Random Forest (RF) performed best on the original ECG data. MLP performed best on the wearable-like ("modified") data.	**Best Metrics (Original ECG - GB Model):** • AUROC: 0.85 • Recall: 0.80 • F1-score: 0.79 **Best Metrics (Wearable-like - MLP Model):** • AUROC: 0.75 • Recall: 0.80 • F1-score: 0.72	• **Context:** Driving stress (Rest, Highway, City). • **Limitations:** "Stress is a subjective mental state"; original dataset used invasive sensors; private dataset had very small sample size (4).
Al-Nafjan et al. [62]	GSR (EDA)	WESAD (Public)	15	• **Features:** Statistical features. • **Classifier:** KNN (k-Nearest Neighbor) performed best (better than SVM and RF). • **Balancing:** Used SMOTE.	**Best Metrics (Autoencoder + KNN):** • Accuracy: 96.90% • Precision: 96.90% • Recall: 96.90% • F1-score: 96.90%	• **Context:** Binary classification: Stress vs. Non-Stress (Neutral + Amusement). • **Limitations:** Small dataset (15 subjects); Single modality (GSR only); Reliance on SMOTE for balancing (potential synthetic bias).
Park et al. [10]	**A:** HR (Heart Rate), GSR (Galvanic Skin Response). **B:** Voice recordings (88 eGeMAPS features). **C:** Combination of both.	Private (SAD Patients)	25	• **Features:** Statistical features (A) + eGeMAPS (B). • **Classifiers:** CatBoost and LightGBM performed best. Also tested Random Forest and XGBoost.	**Best Metrics (CATBoost):** • **K-SPS - Social Phobia Scale:** AUROC: 0.852 • **PERS - Rumination:** AUROC: 0.866 **A Only (LightGBM for STAI-Trait):** AUROC: 0.819	• **Context:** VR Exposure Therapy (Self-intro task). Predicting "Severe" (top tertile) vs "Non-severe" symptom groups. • **Limitations:** Small sample size (25); specific demographic (young, educated Koreans); potential overfitting due to small data; lack of HRV analysis (HR sampled <100Hz).
Ihmig et al. [43]	ECG, EDA, RSP	Private (Spider-fearful)	57	• **Features:** Statistical features (time domain). • **Classifier:** Bagged Trees (Ensemble) performed best.	**Best Metrics (2-Level → Bagged Trees + SB labeling):** • Accuracy: 89.8% • Kappa: 0.78 **Best Metrics (3-Level → Bagged Trees + SB labeling):** • Accuracy: 74.4% • Kappa: 0.59	• **Context:** Spider-phobia exposure (VRET). Real-time detection (10s window) for "closed-loop" biofeedback. • **Limitations:** Subjective labels may contain noise; RSP signals found not useful; single specific phobia type.

Table Note: In this table, "**Features**" refers to the specific physiological, statistical, or engineered variables extracted from the raw signals before classification. "**Best Metrics**" denotes the highest performance scores achieved by the optimal model configuration reported in the respective study.

Table 4: Comparative performance of selected anxiety and stress detection studies that used deep learning approaches.

Study (Ref)	Data Modality(s)	Dataset	# Subjects	Model Architecture	Key Performance (Metric)	Clinical Context / Limitations
Deep Learning Approaches						
Ahmed et al. [65]	EEG	Private	945	• **Features:** PSD and Functional Connectivity (FC). • **Classifiers:** Compared ANN, KNN, LSTM, Bi-LSTM, and CNN-LSTM.	**Best Metrics:** • 98.94% Accuracy for Acute Stress Disorder (using KNN or LSTM on specific band features). • 96.83% Accuracy for Adjustment Disorder (CNN-LSTM).	• **Context:** Diagnosis of 6 main and 9 specific psychiatric disorders (including Anxiety, OCD, PTSD, Adjustment Disorder). • **Limitations:** EEG data collected from finite subjects; potential comorbidity overlap not fully captured; controlled environment (not real-world).
Roy et al. [49]	EEG	STEW (Public)	48	• **Preprocessing:** DWT (Discrete Wavelet Transform) for decomposition/denoising. • **Features:** CNN (1D). • **Classifier:** Hybrid CBGG (CNN + BiLSTM + 2 GRU layers). • **Comparison:** Compared against CNN-RNN, CNN-LSTM, CNN-GRU, CBRR, CBLL.	**Best Metrics (CBGG Model):** • Accuracy: 98.10% • Precision: 98.27% • Sensitivity (Recall): 98.08% • Specificity: 97.76% • F1-Score: 98.20%	• **Context:** Stress vs. Relaxed state classification. • **Limitations:** "Complex" DL model with significant parameter tuning required; high running time.
Kim et al. [81]	EEG, GSR	Private	30	• **Features:** Spectrograms (Grayscale images from time series). • **Classifiers:** Compared 5 CNNs (ResNet-50, ResNet-152, EfficientNet-b0, DenseNet-161, Inception-v3) and a Vision Transformer (ViT). • **Best Model:** Multi-column ResNet-50 (combining EEG and GSR features).	**Best Metrics (Multi-column ResNet-50):** • AUROC: 0.954 • Accuracy: 86.5% • Recall: 88.7% • Precision: 92.8% • F1-Score: 0.898	• **Context:** VR-based mock interview to induce social stress. "Daily life" use-case with unobtrusive sensors. • **Limitations:** Small dataset (30 subjects); only 1-channel EEG used (less spatial info); specific stressor (interview) may not generalize to all daily stress.
Tanwar et al. [80]	ECG, EMG, EDA, etc.	WESAD (Public)	15	**Features/Classification:** Hybrid CNN-LSTM (2D CNN layers followed by LSTM layers).	**Best Metrics (Stress Class):** • Accuracy (Overall): 90.20% • Precision: 90.57% • Recall: 90.39% • F1-score: 90.39%	• **Context:** 3-class classification (Baseline, Stress, Amusement). Used chest-worn data (RespiBAN) as it provided better results than wrist-worn. • **Limitations:** Small dataset (15 subjects); computational complexity of deep learning models.
Kang et al. [72]	ECG	ST Change Database, WESAD	15 (ST Change), 15 (WESAD)	• **Features:** Spectrograms (Time & Frequency domain images). • **Classifier:** Ensemble of CNN-LSTM models.	**Best Metrics (Frequency Domain Ensemble):** • Accuracy: 98.3% • Sensitivity (Recall): 100% • Specificity: 96.7% • Precision: 96.6%	• **Context:** Stress vs. Non-stress classification using ECG spectrograms. • **Limitations:** Data augmentation used to increase sample size (potential overfitting risk); complex preprocessing (FFT/Spectrogram conversion).
Masuda et al. [70]	EEG, EOG, EMG, etc.	DEAP (Public)	32	**Features/Classification:** Multi-Input CNN-LSTM	**Best Metrics (Multi-Input CNN-LSTM, 4-level Fear):** • Accuracy: 98.79% • F1-Score: 99.01% (Acc)	• **Context:** Classifying Fear levels (No, Low, Medium, High). Fear defined by VAD quadrants (Low Valence, High Arousal, Low Dominance). • **Limitations:** "Subjective labeling" (relied on self-reports); DEAP is not specific to fear (music videos); performance dropped significantly in user-independent (LOSO) validation without calibration.
Behinaein et al. [83]	ECG	WESAD (Public), SWELL-KW	15 (WESAD), 25 (SWELL-KW)	**Features/Classification:** End-to-end CNN + Transformer Encoder.	**Best Metrics (WESAD, LOSO with 10% Fine-Tuning):** • Accuracy: 91.1% • F1-Score: 83.3% **Best Metrics (SWELL-KW, LOSO with 10% Fine-Tuning):** • Accuracy: 71.6% • F1-Score: 74.2%	• **Context:** Stress vs. Non-stress. End-to-end learning from raw ECG. • **Limitations:** LOSO performance without fine-tuning was lower (80.4% Acc); requires subject-specific calibration (fine-tuning) to achieve SOTA results.
Arjun et al. [84]	EEG	DEAP (Public)	32	**Features/Classification:** Vision Transformer (ViT)	**Best Metrics (Raw EEG, 6s window):** • Valence Accuracy: 99.4% • Arousal Accuracy: 99.1% **Best Metrics (CWT Images, 6s window):** • Valence Accuracy: 97% • Arousal Accuracy: 95.75%	• **Context:** 2-class classification (High/Low Valence, High/Low Arousal). "Raw signal approach surprisingly performs much better than CWT." • **Limitations:** ViT requires large data/resources; DEAP is not specific to clinical anxiety (music videos); results are on very short windows (6s) which may boost accuracy but have lower clinical interpretability.

Table Note: In this table, "**Features**" refers to the specific physiological, statistical, or engineered variables extracted from the raw signals prior to classification. "**Best Metrics**" denotes the highest performance scores achieved by the optimal model configuration reported in the respective study.

Table 5: Comparative performance of selected anxiety and stress detection studies that used deep learning approaches- continued

Study (Ref)	Data Modality(s)	Dataset	# Subjects	Model Architecture	Key Performance (Metric)	Clinical Context / Limitations
Deep Learning Approaches						
Guo et al. [87]	EEG	DEAP, DREAMER	32 (DEAP), 23 (DREAMER)	**Features/Classification:** SAG-CET (Cascade Scale-aware Adaptive Graph Convolutional Network and Cross-EEG Transformer)	**Best Metrics (DEAP (Subject-Dependent)):** • Valence Accuracy: 98.89% • Arousal Accuracy: 98.92% **Best Metrics (DREAMER (Subject-Dependent)):** • Valence Accuracy: 99.08% • Arousal Accuracy: 99.21%	• **Context:** Emotion recognition (Valence/Arousal). Investigates multiscale (temporal) and multichannel (spatial) interactions. • **Limitations:** Experiments were subject-dependent (trained and tested on the same subjects), so generalization to new subjects is unknown; binary classification only.
Al-Nafjan et al. [62]	GSR (EDA)	WESAD (Public)	15	• **Features:** 14-layer Autoencoder (Automatic features). • **Classifier:** KNN (k-Nearest Neighbor) performed best (better than SVM and RF). • **Balancing:** Used SMOTE.	**Best Metrics (Autoencoder + KNN):** • Accuracy: 98.20% • Precision: 98.20% • Recall: 98.20% • F1-score: 98.20%	• **Context:** Binary classification: Stress vs. Non-Stress (Neutral + Amusement). • **Limitations:** Small dataset (15 subjects); Single modality (GSR only); Reliance on SMOTE for balancing (potential synthetic bias).
Joshi et al. [71]	EEG	DASPS, Private	23 (DASPS), 50 (Private)	• **Features:** Higher Order Statistics (HOS) - specifically Third Order Cumulants (TOC). • **Classifier:** BiLSTM	**Best Metrics:** • **DASPS Database:** Accuracy: 87%. • **Private Database:** Accuracy: 98%	• **Context:** Exam stress (Pre-exam vs Post-exam). Single channel EEG (Fp1). • **Limitations:** Single channel may lack spatial resolution; limited physiological signals (only EEG); private dataset demographics limited to students (15-20 years).
Ebrahimpour et al. [78]	ECG	Private	40	• **Preprocessing:** Continuous Wavelet Transform (CWT) to generate Scalograms (Time-Frequency images). • **Classifier:** ResNet50	**Best Metrics (ResNet50, 5s segments):** • Accuracy: 94.92% • AUC: 0.991 • MCC: 0.8983 • F1-Score: 94.89% • Precision: 95.45% • Recall: 94.33%	• **Context:** PTSD diagnosis (DSM-5 criteria). Time-frequency analysis of ECG. • **Limitations:** Small sample size (40); retrospective study; binary classification only (did not distinguish from other comorbidities like depression).
Al-Ezzi et al. [66]	EEG	Private	88	• **Features:** Effective Connectivity (EC) using Partial Directed Coherence (PDC) from DMN regions. • **Classifier:** CNN + LSTM	**Best Metrics (CNN+LSTM for 3-class SAD + HC):** • Average Accuracy: 93% • Specific Class Accuracies: Severe (92.86%), Moderate (92.86%), Mild (96.43%), HC (89.29%). • Average Sensitivity: 86% • Average Specificity: 95% • Average Precision: 86%	• **Context:** Resting-state EEG. Classifying severity (Mild, Moderate, Severe) vs Control. Focus on Default Mode Network (DMN). • **Limitations:** Small input feature size led to overfitting (addressed with early stopping); high density EEG recommended for future.
Ghonchi et al. (2024) [73]	EEG	DASPS (Public)	23	**Features/Classifier:** Spatio-temporal conversion (11x11 scalp maps) + Hybrid CNN-LSTM + SE Attention.	**Best Metrics:** • **2-Class (Binary - Normal vs Anxious):** Accuracy: 94.24% (200ms window). • **4-Class (Normal, Light, Moderate, Severe):** Accuracy: 92.58% (200ms window).	• **Context:** Anxiety detection using DASPS. Scalp map generation allows preserving spatial info. • **Limitations:** Small dataset; class imbalance in 4-class scenario; sliding window size affects performance (200ms optimal).
Ghonchi et al. (2025) [79]	EEG	DASPS (Public)	23	**Classifier:** MVCA-Net (CNN+Attention)	**Best Metrics:** • **2-Class (Binary):** Accuracy: 82.94% • **4-Class:** Accuracy: 74.05% • **Runtime:** Very fast (Epoch time: 8s vs 79s for their previous method [73]).	• **Context:** Focuses only on frequency domain features (unlike their previous spatio-temporal work [73]). • **Limitations:** Lower accuracy than their previous CNN-RNN work [73] (94.24%), but significantly faster/more efficient; relies solely on frequency features, missing spatial/temporal relationships.

Table Note: In this table, "**Features**" refers to the specific physiological, statistical, or engineered variables extracted from the raw signals prior to classification. "**Best Metrics**" denotes the highest performance scores achieved by the optimal model configuration reported in the respective study.

Firstly, the choice of performance metric is crucial. Although accuracy is the most commonly reported metric and most of the researchers rely on this metric, it can be highly misleading in clinical contexts where class imbalance (i.e., fewer anxious than non-anxious samples) is a common problem. A model could achieve 90% accuracy simply by always predicting the majority class. For this reason, metrics like the F1-score, which represents the harmonic mean of precision and recall, and the Area Under the Receiver Operating Characteristic Curve (AUROC or AUC), are more robust indicators of performance. The emphasis placed on F1-score by Vaz et al. [56] and on AUROC by Park et al. [10] and Kim et al. [81] represents a move towards more rigorous and clinically relevant evaluation. Furthermore, recent studies have adopted even more robust metrics for imbalanced clinical data. For instance, Muhammad et al. [19] reported a Kappa score of 0.88 to account for chance agreement, while Ebrahimpour et al. [78] utilized the Matthews Correlation Coefficient (MCC) to validate their high accuracy of 94.92%. The lack of an acceptable set of reporting metrics remains a significant barrier to fair comparison across studies.

Secondly, the validation methods dramatically affect reported outcomes. The highest performances, such as the accuracies above 98% as reported by Kang et al. [72], and Roy et al. [49], are often achieved on smaller, private datasets using a subject-dependent or intra-subject validation model. In this method, the model is trained and tested on data from the same pool of individuals (e.g., using k-fold cross-validation where splits can contain data from the same person in both train and test sets). While this demonstrates that a model can effectively learn an individual's unique physiological patterns, it does not guarantee that it will work on a new, unseen person. In contrast, studies using a more stringent subject-independent validation, such as leave-one-subject-out cross-validation, where the model is tested exclusively on data from individuals not seen during training, tend to report more modest, and likely more realistic, accuracies [83, 56]. This highlights a significant generalizability gap between what is achievable in a personalized setting versus a general population screening tool.

4.2. Performance Across Modalities and the Rise of Multimodal Fusion

Single modality systems can demonstrate high effectiveness in controlled settings. For instance, Al-Nafjan et al. [62] reported over 96% accuracy using only GSR features from the WESAD dataset. However, such systems often lack the robustness necessary to address the complexities of real-world scenarios. A single signal can be contaminated by noise or influenced by non-anxious confounds (e.g., physical activity raising heart rate).

The superiority of multimodal approaches stems from two primary mechanisms, complementarity and robustness, as discussed in recent papers on this topic [11, 15]. Physically, anxiety involves a synchronized activation of both the CNS and ANS. Multimodal fusion captures this full picture. For instance, EEG captures the cognitive state (e.g., worry), whereas EDA reflects the corresponding physiological arousal. This complementary information prevents the model from misinterpreting a high heart rate caused by physical exercise as an anxiety attack. Furthermore, advanced deep learning fusion techniques, particularly those incorporating attention mechanisms, enhance robustness by dynamically learning to weight different modalities [81]. If one signal (e.g., ECG) becomes corrupted by motion artifacts, the network can learn to down-weight its contribution and rely more heavily on stable features from another stream (e.g., EDA), effectively mitigating noise.

Common fusion strategies include early fusion, late fusion, and hybrid fusion. Early fusion concatenates feature vectors from multiple modalities prior to classification. Late fusion aggregates the outputs or probabilities produced by independent classifiers. Hybrid fusion combines multimodal representations at intermediate layers of a deep network [11]. While adding modalities generally improves performance, it is not always the case that "more is better". It should be noted that increasing the sensor numbers introduces higher computational costs, battery drain, and points of failure. Therefore, the goal is to identify the minimal optimal set of modalities that maximizes clinical utility.

Direct comparative studies and systematic surveys provide rigorous support for the superiority of multimodal fusion. For instance, in their comprehensive

review of deep learning approaches, Kyrou et al. [11] concluded that multimodal frameworks consistently outperform unimodal ones across diverse datasets. By combining information from different physiological data, these models create a more holistic and reliable representation of an individual's state.

Another way to categorise fusion models comes from the origin of the signals being used. In peripheral fusion, models combine multiple autonomic nervous system signals, such as ECG, EDA, and respiration. For instance, Vaz et al. [56] conducted a comprehensive analysis on the WESAD dataset and demonstrated that a multimodal model combining ECG, EDA, and EMG achieved a high accuracy of 92.0% and an AUC of 0.980, consistently outperforming single-modality baselines. This reinforces findings by Tanwar et al. [80], confirming that the combination of cardiac and electrodermal data leads to better stress classification performance compared to any single stream alone.

The second category, Central + Peripheral Fusion, represents a more advanced approach, combining brain activity from EEG with peripheral signals. Multiple studies confirm the efficacy of this synergy. The study performed by Kim et al. [81] is a landmark in this field. Their multi-column CNN fusing EEG and GSR data significantly outperformed models trained on either modality alone, achieving an outstanding AUROC of 0.954. Similarly, Guo et al. [90] demonstrated that fusing EEG with eye movement data using nonlinear correlation analysis yielded higher classification accuracy than using EEG features alone. These findings suggest that the ultimate goal for anxiety detection should lie in simultaneously capturing the cognitive state via EEG and the body's arousal response via peripheral signals. However, implementing this approach in real-world settings poses significant challenges. Traditional EEG is limited by lengthy setup times and susceptibility to motion artifacts. The emergence of consumer-grade, low-channel devices such as dry-electrode headbands or in-ear EEG offers a practical pathway for translating this powerful multimodal approach to ambulatory monitoring.

4.3. The Ascendancy of Deep Learning: From Feature-Based to End-to-End Systems

Although traditional machine learning pipelines built on hand-crafted features remain highly accurate and offer greater interpretability [56, 19], the performance ceiling and methodological innovations are increasingly shifting towards deep learning. Rather than a singular shift to end-to-end learning, this transition encompasses a spectrum of approaches that minimize, transform, or entirely eliminate the need for manual feature engineering.

It is crucial to distinguish between these operational frameworks when evaluating performance. Many highly successful models employ feature-based deep learning. For instance, Ahmed et al. [65] applied CNN-LSTMs to Power Spectral Density (PSD) and functional connectivity features, while Al-Ezzi et al. [66] achieved high accuracy using Partial Directed Coherence (PDC) features. A second prevalent approach is representation-based hybrid learning, where raw signals are transformed into structured inputs to exploit specific network strengths. This is evident in hybrid CNN-LSTM architectures [80, 70] that learn from generated representations like ECG spectrograms [72] or EEG scalp maps [73].

Finally, strict end-to-end learning bypasses explicit feature extraction entirely. Advanced architectures, such as Transformer models with their attention mechanisms, have shown the ability to achieve high performance on raw 1D ECG signals with minimal preprocessing (e.g., Behinaein et al. [83]), relying on the network to discover the most discriminative biomarkers directly from the source data.

Furthermore, unsupervised deep learning offers a compelling solution to the feature engineering bottleneck. The work of Al-Nafjan et al. [62] and Shikha et al. [89] demonstrated that Autoencoders can learn more potent feature representations from raw GSR and EEG data than traditional statistical methods. The fact that a simple kNN classifier achieved higher accuracy with these learned features is a powerful testament to their quality. This trend suggests a future in which the reliance on domain-specific signal processing knowledge may be reduced, as models become increasingly capable of discovering the most

discriminative biomarkers directly from the source data.

5. Discussion: Synthesis, Challenges, and Implications

The comprehensive review and comparative analysis of the literature reveal a field of research that is both rapidly advancing and fraught with significant challenges. While the potential for AI-driven anxiety detection is clear, a critical discussion of the key findings and persistent obstacles is necessary to chart a path toward clinical translation.

5.1. Principal Findings: A Synthesis of Anxiety's Biosignatures

Reviewing diverse methodologies and datasets leads to finding a remarkable set of physiological signatures for anxiety.

A remarkable set of physiological signatures emerges from the models we have discussed. These biosignatures, which can both be identified by traditional feature engineering and implicitly learned by deep learning models, are different windows into the same integrated psycho-physiological stress response. They form the foundation of all automated detection efforts.

The most consistent neural correlate of anxiety identified in EEG is a marked shift in spectral power: a decrease in alpha band (8-12 Hz) power, particularly in frontal and prefrontal regions, accompanied by a significant increase in beta band (12-30 Hz) power [16, 17]. A distinct transition from the state of quiet, resting awareness or internal focus to anxiety is reflected by a neurophysiological shift from alpha band oscillation, which facilitates internal focus, to beta oscillation, which underpins cortical arousal and cognitive hypervigilance. Additionally, evidence indicates that other underlying dynamics also contribute to this neurophysiological transition. For instance, Peddi et al. [91] reported that an increase in theta band (4-8 Hz) connectivity correlates with PTSD severity. Their finding suggested that the state transition mechanism is characterized by modulation in network information exchange, rather than alterations in absolute power. Furthermore, changes in frontal alpha asymmetry have been linked to anxiety reduction [91, 92, 93]. This distinct spectral shift explains the his-

torical success of frequency-domain features such as PSD combined with non-linear classifiers. However, the complex topological dependencies—specifically the functional connectivity between brain areas and the asymmetry across hemispheres—are arguably better captured by deep learning models like CNNs and Graph Neural Networks, which can learn these spatial patterns without manual feature extraction [81, 87].

In the context of the heart and circulatory system, as ascertained by means of ECG, one characteristic observation is that the reduction of Heart Rate Variability accompanies anxiety states. HRV is a measure of autonomic nervous system adaptability; lower HRV indicates a less adaptive and physiologically inflexible state. This observation suggests a shift from the restorative parasympathetic ("rest-and-digest") activity toward sympathetic ("fight-or-flight") dominance. This shift constitutes the physiological signature of an organism preparing for a perceived threat. Various conventional measures can quantify the reduction in HRV. Time-domain measures, such as SDNN, defined by the standard deviation of NN intervals, and RMSSD, defined as the root mean square of successive differences, are uniformly found to be lower in individuals with anxiety [94]. SDNN reflects the overall variability, whereas RMSSD primarily indexes short-term, vagally mediated variability. In the frequency domain, the dominance by the sympathetic system is often signaled by an increase in the LF/HF ratio, indicating a rise in sympathetic tone; yet its particular meaning remains the focus of current debate within the academic literature [24]. The temporal evolution of these HRV metrics highlights why RNNs and LSTMs, with their internal memory mechanisms, have consistently outperformed static classifiers in detecting anxiety using cardiac signals, as they can effectively model the time-dependent dynamics of autonomic regulation [80, 67].

EDA has been called a direct and clear measure of sympathetic nervous activity. It offers a very robust biosignature for anxiety-related arousal. The signature is two-fold. First, individuals with anxiety tend to exhibit an elevated tonic SCL, indicating a persistently heightened state of physiological arousal that fails to recover promptly when a stimulus has been removed [24]. Second,

they tend to have more repeated and elevated phasic SCRs. These are transient peaks in skin conductance that correspond to discrete emotional or cognitive events, such as moments of worry, rumination, or reaction to a stressor [95]. Because these phasic bursts are sporadic and event-driven, traditional statistical features may fail to capture their full complexity. This limitation drives the success of unsupervised deep learning methods, such as Autoencoders, which have been shown to learn more discriminative representations from raw EDA signals compared to hand-crafted features [82].

In summary, findings from EEG, ECG, and EDA represent interlocking components of a unified physiological response. An anxious thought or perception, reflected in elevated EEG beta power, triggers a coordinated arousal command from the brain that activates the sympathetic nervous system. This in turn drives the heart into a faster, more regular rhythm, triggering a reduced HRV, while simultaneously activating the sweat glands, as reflected by increased EDA. The most successful AI models, particularly multimodal fusion techniques, are implicitly or explicitly learning to detect this synchronized, multi-system physiological signature of anxiety.

5.2. Grand Challenges and Open Questions

Despite promising results from laboratory settings, the translation of these AI models into effective and trusted real-world clinical tools is hindered by several major and unsolved challenges.

First and foremost are dataset limitations and the generalizability problem, which is further exacerbated by the field's growing shift from controlled laboratory protocols to ambulatory, in-the-wild contexts. As highlighted in Tables 3 and 4, the vast majority of current research relies on public datasets such as WESAD or DEAP, which are collected in highly controlled environments, apply standardized stress elicitation paradigms, and typically involve fewer than 50 healthy participants. However, real-world anxiety monitoring involves physical movement and unscripted interactions, introducing significant noise into neural and physiological recording. In particular, motion artifacts in ECG and PPG

signals and electrode instability in EEG severely degrade model performance outside controlled laboratory settings [7, 11]. Consequently, models trained on clean, stationary data frequently fail to generalize to the noisy and dynamic conditions encountered in daily life. This generalizability gap remains a critical barrier to clinical utility. Indeed, this challenge is not unique to anxiety; a comprehensive meta-analysis on major depressive disorder also found evidence of publication bias and a negative correlation between sample size and reported accuracy, suggesting that findings from smaller, controlled studies may not generalize to larger, more heterogeneous clinical populations [96].

A second major obstacle concerns confounding variables and specificity. A core challenge is that the physiological signatures of anxiety are not unique. The body's arousal response is generic. Physical exercise, drinking caffeine, experiencing positive excitement, or engaging in high cognitive load can all lead to increased heart rate, higher beta brainwave activity, and increased sweating. In these cases, signals are indistinguishable from an anxiety response when viewed in isolation. Furthermore, the context of the stress-eliciting stimuli significantly modulates the physiological response. Prior research indicates that social evaluative threats (e.g., public speaking in the Trier Social Stress Test) tend to elicit distinct and often more intense autonomic patterns compared to purely cognitive challenges (e.g., the Stroop test) [15]. This creates a high risk of false positives, which could lead to unnecessary user distress or inappropriate interventions. Therefore, a grand challenge for the field is distinguishing anxiety-related responses from non-emotional arousal. How can models learn to differentiate between a panic attack and jogging up a flight of stairs? The integration of contextual sensors, such as accelerometers to detect motion or GPS to understand a user's environment, is a promising direction, but sophisticated and reliable sensor fusion techniques are still in their infancy.

A third drawback is the black box problem, which describes the fact that models or feature engineering pipelines are not explainable, resulting in a need for Explainable AI (XAI). Additionally, despite gains in accuracy, increasing complexity in deep learning models is accompanied by a loss of interpretabil-

ity. For a technology to be adopted in a high-stakes domain like healthcare, it must be trustworthy, and trust requires understanding. A clinician will not, and should not prescribe treatment or alter a care plan based solely on an AI's output without understanding the rationale behind its application. For example, was the anxiety prediction based on a subtle change in HRV, a spike in EDA, or a specific EEG pattern? Such justification is essential for clinical decision-making, accountability, and troubleshooting erroneous predictions. The integration of Explainable AI techniques is therefore not optional but a necessity for clinical translation. Methods such as SHAP (Shapley Additive exPlanations) or attention maps, which can highlight the features or signal segments most influential in a given prediction, must become a standard part of the model development and validation process. Furthermore, developing computationally efficient architectures, such as the frequency-focused MVCA-Net, is also crucial, as it reduces the computational resources required for training and allows for faster iteration on developing and evaluating XAI capabilities [79].

Finally, the development of technology capable of inferring mental states raises substantial ethical considerations that must be addressed proactively. These responsibilities include: *i)* data privacy and security, which represent a central ethical concern, particularly with respect to how highly sensitive personal health data are stored, protected, used, and de-identified. The risk of misuse and unauthorized access is significant and must be carefully mitigated; *ii*) algorithmic bias, whereby insufficient demographic diversity in training datasets may lead to a reduced model accuracy for underrepresented groups, potentially exacerbating existing health disparities; and *iii*) iatrogenic effects, wherein constantly monitoring by an anxiety detector may paradoxically increase a user's anxiety, leading to a harmful feedback loop of hypervigilance and misinterpretation of normal physiological variations. Developing clear principles for responsible AI in mental health is a critical parallel research trajectory that must evolve alongside the technology itself.

5.3. From Classification to Comprehension: The Path Toward Personalized Intervention

The vast majority of studies reviewed in this paper, from traditional SVMs to advanced Transformers, focus on the core technical task of classification, referring to detecting the presence or severity of anxiety within a discrete, short time window. While demonstrating the feasibility of identifying anxiety's biosignatures is a crucial initial step, a purely classification-centric approach does not fully address the complex demands of clinical practice, which requires a deeper comprehension of an individual's condition. The future of this field lies in moving beyond simple detection toward a more holistic understanding.

A key limitation of many current models is their inability to differentiate between the physiological manifestations of different anxiety disorders or their varying levels of severity. For example, are the explosive, acute physiological signatures of a panic attack fundamentally different from the chronic, free-floating hyperarousal of GAD? While this remains an open question, emerging research is increasingly demonstrating the feasibility of anxiety subtyping and severity grading. The ability of some models, such as the spatio-temporal CNN-RNN proposed by Ghonchi et al. [73], to achieve accuracy above 92% in distinguishing between four different levels of anxiety on the DASPS dataset represents a foundational step in this direction. This is further supported by the work of Luo et al. [46], who successfully classified GAD into four severity levels, including healthy, mild, moderate, and severe, using a CatBoost model trained on EEG connectivity features, and by the work of Al-Ezzi et al. [82], who used a hybrid CNN-LSTM to distinguish between four levels of SAD with 92.86% accuracy. These studies demonstrate that discernible, classifiable patterns exist not only between anxiety and non-anxiety states but also across different levels of symptom severity.

The ultimate goal of this progression is to develop a rich, longitudinal digital phenotype of a patient's anxiety. Such a phenotype would provide a dynamic, moment-by-moment quantification of an individual's condition, derived from multimodal sensor data collected over an extended period of weeks or months.

Such a profile would track not only the presence of anxiety but also its intensity, duration, frequency, and environmental context, offering clinicians a unique, objective insight into the patient's lived experience between traditional clinical appointments. To realize this vision, the field must embrace personalized modeling approaches, such as subject-specific domain adaptation. These techniques calibrate algorithms to an individual's unique physiological baseline, overcoming the limitations of one-size-fits-all population models that often fail to account for inter-subject variability [12].

Deep insight into this topic is a necessary step toward unlocking this technology's true potential, enabling person-centered, just-in-time adaptation interventions (JITAIs). An entity with knowledge of a person's particular anxiety profile might switch from passive monitoring to active management. For instance, it might induce a brief mindfulness exercise immediately before an anticipated stressful event, such as giving a public speech, offer real-time biofeedback to regulate breathing during periods of heightened physiological arousal, or recommend therapeutic support if a prolonged anxious state is detected. The above is representative of the ultimate summit of clinical application in the field—not simply to detect anxiety, but also to use that detection to actively and individually manage its effects. In this manner, the rich, complicated world of AI-based anxiety detection can be seen as an interplay of detecting key biosignatures, new computation methods, significant impediments that thwart progress, and future directions that frame the discipline. The four key parts of this world are schematically shown in Figure 1, which provides a conceptual map of the current status and future trajectory of this research field.

6. Conclusion and Future Outlook

This systematic review has charted the landscape of artificial intelligence for anxiety detection, highlighting a vibrant and rapidly evolving field that sits at the critical intersection of computational science, physiological sensing, and clinical need. The research overwhelmingly confirms that anxiety manifests

Figure 1: A synthesis of the landscape of AI for anxiety detection. This figure summarizes four core pillars of the field reviewed in this paper. It illustrates the primary data modalities and their key physiological biomarkers, the main computational methods used for analysis, the grand challenges that currently limit clinical translation, and the key future directions required to move the field forward.

through a consistent and measurable set of biosignatures spanning the central and autonomic nervous systems, including changes in EEG spectral power, reductions in heart rate variability, and increases in electrodermal activity. AI models, particularly deep learning architectures like hybrid CNN-LSTMs and Transformers, have proven remarkably effective at detecting these signatures in controlled settings, often with high accuracy.

However, despite these technical successes, the field is at a crucial inflection point. The translation from laboratory-based proof-of-concept to real-world clinical application is hampered by significant challenges, most notably the lack of large-scale, clinically-validated datasets, the problem of confounding physiological states, and the black box nature of complex models. The promise of this technology will only be realized if the research community pivots to address these hurdles directly. To that end, the future of the field should be guided by three key directions as outlined below.

1. **Methodological Innovation in Data Science and Generalization:** current models often lack robustness when moved from one scenario to another. To address this, future algorithms must demonstrate cross-context generalization which is the ability of detecting anxiety across diverse contexts (e.g., applying a model trained on public speaking to a driving scenario). The success of this endeavor will also yield insights into the key, consistent features of anxiety, fostering a productive bidirectional exchange between computational and psychological science. Researchers should prioritize computational techniques like Domain Adaptation and Transfer Learning to bridge the generalization gap. Furthermore, the limitations of centralized datasets must be overcome through Federated Learning, allowing models to train on multi-institutional data without compromising privacy, alongside sophisticated multimodal fusion to intelligently weigh data streams against environmental noise.

2. **A Shift Towards Efficient, Longitudinal Monitoring:** The focus must evolve from detecting short-term anxious states to longitudinal moni-

toring that captures the dynamic nature of anxiety over weeks and months. The goal is to construct a digital phenotype for each individual. However, continuous monitoring generates massive data streams that strain wearable batteries. To address the challenge of computational efficiency, future research must prioritize the development of lightweight model architectures (e.g., TinyML, model compression, or quantized neural networks). These techniques allow complex deep learning models to run directly on low-power edge devices ("on-chip"), enabling real-time, privacy-preserving inference without the latency of cloud offloading.

3. **Rigorous Clinical Validation and Integration:** To overcome the black box trust barrier, the field must prioritize Explainable AI alongside formal clinical validation. Success is measured not just by accuracy, but by impact on patient outcomes. Future work must move beyond benchmark datasets to conduct Randomized Controlled Trials (RCTs) that prove these systems can improve diagnostic accuracy and treatment efficacy in real-world settings. Additionally, researchers must focus on the human-computer interaction aspects, ensuring AI tools provide interpretable insights that support, rather than replace, the clinical judgment of healthcare professionals.

In conclusion, AI-driven anxiety detection is no longer a theoretical possibility but an emerging technological reality. By focusing on building robust, generalizable, explainable, and clinically validated systems, the research community has the potential to transform anxiety assessment from a subjective, episodic art into an objective, continuous science, ultimately providing profound benefits to the millions of individuals affected by these debilitating conditions.

Appendix A. Summary of common anxiety questionnaires

Table A.6: Summary of Common Anxiety Questionnaires

Questionnaire	**Items**	**Approx. Time**	**Focus**
Generalized Anxiety Disorder-7 (GAD-7)	7	2-3 minutes	Screening and severity of generalized anxiety symptoms
Hamilton Anxiety Rating Scale (HAM-A)	14	10-15 minutes	Clinician-rated assessment of anxiety symptom severity
Beck Anxiety Inventory (BAI)	21	5-10 minutes	Intensity of somatic and cognitive anxiety symptoms
State-Trait Anxiety Inventory (STAI)	40 (20 state, 20 trait)	10-15 minutes	Distinguishes between current (state) and general (trait) anxiety
Anxiety Symptoms Questionnaire (ASQ)	17	5-10 minutes	Frequency and intensity of anxiety symptoms
Social Anxiety Disorder Severity Measure (Adults)	10	3-5 minutes	Severity and monitoring of social anxiety disorder
Zung Self-Rating Anxiety Scale (SAS)	20	5-10 minutes	Self-perceived anxiety levels, including somatic and cognitive symptoms
Depression Anxiety Stress Scales (DASS-21/DASS-42)	21 or 42	5–10 (DASS-21) or 10–15 (DASS-42) minutes	Depression, anxiety, and stress (multidimensional)
Personal Report of Public Speaking Anxiety (PRPSA)	34	10-15 minutes	Anxiety specifically related to public speaking
Personal Report of Communication Apprehension – Public Speaking Subscale (PRCA-PS)	24	5-10 minutes	Apprehension and anxiety in public speaking contexts
Public Speaking Anxiety Scale (PSAS)	17	5-10 minutes	Cognitive, behavioral, and physiological aspects of speech anxiety
Self-Statements During Public Speaking Scale (SSPS)	10	2-3 minutes	Positive and negative self-talk during public speaking
Liebowitz Social Anxiety Scale (LSAS)	24 (plus 24 for avoidance)	10-15 minutes	Social and performance anxiety, including public speaking
Personal Report of Confidence as a Speaker (PRCS)	30	10-15 minutes	Confidence and self-efficacy in public speaking

References

[1] What is anxiety? — mentalhealth.org.uk, Available at: https://www.mentalhealth.org.uk/our-work/public-engagement/mental-health-awareness-week/anxiety-report/what-anxiety, (Accessed 26 March 2025).

[2] Anxiety disorders — who.int, https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders, (Accessed 07 April 2025).

[3] B. Bandelow, S. Michaelis, Epidemiology of anxiety disorders in the 21st century, Dialogues in clinical neuroscience 17 (3) (2015) 327–335.

[4] How accurate are GAD-7 and GAD-2 questionnaires for detecting anxiety disorders? — cochrane.org, https://www.cochrane.org/CD015455/CENTRALED_how-accurate-are-gad-7-and-gad-2-questionnaires-detecting-anxiety-disorders, (Accessed 07 April 2025).

[5] V. L. Clark, J. A. Kruse, Clinical methods: the history, physical, and laboratory examinations, Jama 264 (21) (1990) 2808–2809.

[6] R. L. Spitzer, K. Kroenke, J. B. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: the gad-7, Archives of internal medicine 166 (10) (2006) 1092–1097.

[7] A. Abd-Alrazaq, R. AlSaad, M. Harfouche, S. Aziz, A. Ahmed, R. Damseh, J. Sheikh, Wearable artificial intelligence for detecting anxiety: systematic review and meta-analysis, Journal of medical Internet research 25 (2023) e48754.

[8] S. S. Khalil, N. S. Tawfik, M. Spruit, Exploring the potential of federated learning in mental health research: a systematic literature review, Applied Intelligence 54 (2) (2024) 1619–1636.

[9] A. Halbig, M. E. Latoschik, A systematic review of physiological measurements, factors, methods, and applications in virtual reality, Frontiers in Virtual Reality 2 (2021) 694567.

[10] J.-H. Park, Y.-B. Shin, D. Jung, J.-W. Hur, S. P. Pack, H.-J. Lee, H. Lee, C.-H. Cho, Machine learning prediction of anxiety symptoms in social anxiety disorder: utilizing multimodal data from virtual reality sessions, Frontiers in Psychiatry 15 (2025) 1504190.

[11] M. Kyrou, I. Kompatsiaris, P. C. Petrantonakis, Deep learning approaches for stress detection: A survey, IEEE Transactions on Affective Computing 16 (2) (2024) 499–517.

[12] M.-H. Tayarani-N., S. I. Shahid, Detecting anxiety via machine learning algorithms: A literature review, IEEE Transactions on Emerging Topics in Computational Intelligence 9 (4) (2025) 2634–2657. doi:10.1109/TETCI.2025.3543307.

[13] A. Alaaeldin, R. Alhajj, Ai-based tools for early detection and accurate diagnosis of anxiety disorder (systematic review), Network Modeling Analysis in Health Informatics and Bioinformatics 14 (1) (2025) 1–11.

[14] A.-F. A. Mentis, D. Lee, P. Roussos, Applications of artificial intelligence-machine learning for detection of stress: a critical overview, Molecular Psychiatry 29 (6) (2024) 1882–1894.

[15] M. Khandelwal, A. Sharma, Machine learning and deep learning techniques to detect mental stress using various physiological signals: A critical insight, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 15 (3) (2025) e70035.

[16] R. K. Kanna, S. V. Athawale, M. Y. Naniwadekar, C. Choudhari, N. R. Talhar, S. Dhengre, Anxiety controlling application using eeg neurofeedback system., EAI Endorsed Transactions on Pervasive Health & Technology 10 (1) (2024).

[17] F. Boby, M. T. Z. Lima, N. Zaman, S. Afroz, N. M. Morshed, S. Ferdousi, Brain electrical activity in newly diagnosed generalized anxiety disorder:

A quantitative eeg analysis, Journal of Bangladesh Society of Physiologist 18 (2) (2023) 78–88.

[18] R. Chandrasekar, M. R. Hasan, S. Ghosh, T. Gedeon, M. Z. Hossain, Machine learning to detect anxiety disorders from error-related negativity and eeg signals, arXiv preprint arXiv:2410.00028 (2024).

[19] F. Muhammad, S. Al-Ahmadi, Human state anxiety classification framework using eeg signals in response to exposure therapy, Plos one 17 (3) (2022) e0265679.

[20] F. Demedts, J. Cornelis, B. Reynvoet, D. Sasanguie, F. Depaepe, Measuring math anxiety through self-reports and physiological data, Journal of Numerical Cognition 9 (3) (2023) 380–397.

[21] E. I. Alshanskaia, N. A. Zhozhikashvili, I. S. Polikanova, O. V. Martynova, Heart rate response to cognitive load as a marker of depression and increased anxiety, Frontiers in Psychiatry 15 (2024) 1355846.

[22] M. Adheena, N. Sindhu, S. Jerritta, Physiological detection of anxiety, in: 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), IEEE, 2018, pp. 1–5.

[23] Y. Zheng, T. C. Wong, B. H. Leung, C. C. Poon, Unobtrusive and multimodal wearable sensing to quantify anxiety, IEEE Sensors Journal 16 (10) (2016) 3689–3696.

[24] E. Zhou, M. Soleymani, M. J. Matarić, Investigating the generalizability of physiological characteristics of anxiety, in: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2023, pp. 4848–4855.

[25] L. N. Lyzwinski, M. Elgendi, C. Menon, The use of photoplethysmography in the assessment of mental health: scoping review, JMIR Mental Health 10 (2023) e40163.

[26] A. Baghdadi, Y. Aribi, R. Fourati, N. Halouani, P. Siarry, A. M. Alimi, Dasps: A database for anxious states based on a psychological stimulation, arXiv preprint arXiv:1901.02942 (2019).

[27] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven, Introducing wesad, a multimodal dataset for wearable stress and affect detection, in: Proceedings of the 20th ACM international conference on multimodal interaction, 2018, pp. 400–408.

[28] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: A database for emotion analysis; using physiological signals, IEEE transactions on affective computing 3 (1) (2011) 18–31.

[29] W. Liu, J.-L. Qiu, W.-L. Zheng, B.-L. Lu, Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition, IEEE Transactions on Cognitive and Developmental Systems 14 (2) (2021) 715–729.

[30] A. C. Almeida, F. Canais, R. Maçorano, M. Lopes, H. A. Ferreira, S. Brás, Anxiecg-ppg database: Synchronized ecg and mobile ppg signals during baseline, physical activated, and psychological activated conditions (2024). doi:10.21227/ynse-te33.
URL https://dx.doi.org/10.21227/ynse-te33

[31] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. H. Behzadan, T. Chaspari, Exploring individual differences of public speaking anxiety in real-life and virtual presentations, IEEE Transactions on Affective Computing 13 (3) (2020) 1168–1182.

[32] J. Pešán, V. Jurık, M. Karafiát, et al., Besst dataset: A multimodal resource for speech-based stress detection and analysis, Evaluation 1 (2) (2024) 00.

[33] M. Majid, A. Rauf Butt, I. Fareed Nizami, A. Arsalan, J. Ryu, Proper: Personality recognition based on public speaking using electroencephalography recordings, IEEE Access 13 (2025) 125570–125586. doi:10.1109/ACCESS.2024.3395434.

[34] R. Ghosh, N. Deb, K. Sengupta, A. Phukan, N. Choudhury, S. Kashyap, S. Phadikar, R. Saha, P. Das, N. Sinha, et al., Sam 40: Dataset of 40 subject eeg recordings to monitor the induced-stress while performing stroop color-word test, arithmetic task, and mirror image recognition task, Data in Brief 40 (2022) 107772.

[35] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, O. Shpenkov, Electroencephalograms during mental arithmetic task performance, Data 4 (1) (2019) 14.

[36] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, circulation 101 (23) (2000) e215–e220.

[37] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, L. Albera, Emotion recognition based on high-resolution eeg recordings and reconstructed brain sources, IEEE Transactions on Affective Computing 11 (2) (2017) 244–257.

[38] H. Cai, Z. Yuan, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, et al., A multi-modal open dataset for mental-disorder analysis, Scientific Data 9 (1) (2022) 178.

[39] K. L. Carpenter, P. Sprechmann, R. Calderbank, G. Sapiro, H. L. Egger, Quantifying risk for anxiety disorders in preschool children: a machine learning approach, PloS one 11 (11) (2016) e0165524.

[40] H. Senaratne, L. Kuhlmann, K. Ellis, G. Melvin, S. Oviatt, A multimodal dataset and evaluation for feature estimators of temporal phases of anxiety, in: Proceedings of the 2021 International Conference on Multimodal Interaction, 2021, pp. 52–61.

[41] E. Turcan, K. McKeown, Dreaddit: A Reddit dataset for stress analysis in social media, in: Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019), Association for Computational Linguistics, Hong Kong, 2019, pp. 97–107. doi:10.18653/v1/D19-6213.
URL https://aclanthology.org/D19-6213/

[42] L. Q. Aryan Rastogi, E. Cambria, Stress detection from social media articles: New dataset benchmark and analytical study, in: 2022 IEEE World Conference of Computational Intelligence (WCCI)., 2022.

[43] F. R. Ihmig, F. Neurohr-Parakenings, S. K. Schäfer, J. Lass-Hennemann, T. Michael, On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals, Plos one 15 (6) (2020) e0231517.

[44] J. Tarrant, J. Viczko, H. Cope, Virtual reality for anxiety reduction demonstrated by quantitative eeg: a pilot study, Frontiers in psychology 9 (2018) 1280.

[45] E. Mörtberg, M. Jansson-Fröjmark, A. Pettersson, T. Hennlid-Oredsson, Psychometric properties of the personal report of public speaking anxiety (prpsa) in a sample of university students in sweden, International journal of cognitive therapy 11 (2018) 421–433.

[46] X. Luo, B. Zhou, J. Fang, Y. Cherif-Riahi, G. Li, X. Shen, Integrating eeg and ensemble learning for accurate grading and quantification of generalized anxiety disorder: A novel diagnostic approach, Diagnostics 14 (11) (2024) 1122.

[47] A. Al-Ezzi, N. Kamel, A. A. Al-Shargabi, F. Al-Shargie, A. Al-Shargabi, N. Yahya, M. I. Al-Hiyali, Machine learning for the detection of social anxiety disorder using effective connectivity and graph theory measures, Frontiers in psychiatry 14 (2023) 1155812.

[48] M. Aldayel, A. Al-Nafjan, A comprehensive exploration of machine learning techniques for eeg-based anxiety detection, PeerJ Computer Science 10 (2024) e1829.

[49] B. Roy, L. Malviya, R. Kumar, S. Mal, A. Kumar, T. Bhowmik, J. W. Hu, Hybrid deep learning approach for stress detection using decomposed eeg signals, Diagnostics 13 (11) (2023) 1936.

[50] Y.-Y. Tsai, Y.-J. Chen, Y.-F. Lin, F.-C. Hsiao, C.-H. Hsu, L.-D. Liao, Photoplethysmography-based hrv analysis and machine learning for real-time stress quantification in mental health applications, APL bioengineering 9 (2) (2025).

[51] K. M. Dalmeida, G. L. Masala, Hrv features as viable physiological markers for stress detection using wearable devices, Sensors 21 (8) (2021) 2873.

[52] N. K. Iyortsuun, S.-H. Kim, M. Jhon, H.-J. Yang, S. Pant, A review of machine learning and deep learning approaches on mental health diagnosis, Healthcare 11 (3) (2023).
URL https://www.mdpi.com/2227-9032/11/3/285

[53] H. Ghonchi, S. Ferdowsi, V. Abolghasemi, Common spatial pattern with deep learning for fetal heart rate monitoring, in: 2022 IEEE Workshop on Signal Processing Systems (SiPS), IEEE, 2022, pp. 1–6.

[54] D. Perpetuini, A. M. Chiarelli, D. Cardone, C. Filippini, S. Rinella, S. Massimino, F. Bianco, V. Bucciarelli, V. Vinciguerra, P. Fallica, et al., Prediction of state anxiety by machine learning applied to photoplethysmography data, PeerJ 9 (2021) e10448.

[55] A. Alkurdi, M. He, J. Cerna, J. Clore, R. Sowers, E. T. Hsiao-Wecksler, M. E. Hernandez, Extending anxiety detection from multimodal wearables in controlled conditions to real-world environments, Sensors 25 (4) (2025) 1241.

[56] M. Vaz, T. Summavielle, R. Sebastião, R. P. Ribeiro, Multimodal classification of anxiety based on physiological signals, Applied Sciences 13 (11) (2023) 6368.

[57] S. Sinche, J. Acán, P. Hidalgo, Anxiety detection using consumer heart rate sensors, Engineering Proceedings 77 (1) (2024) 10.

[58] M. Kang, S. Shin, G. Zhang, J. Jung, Y. T. Kim, Mental stress classification based on a support vector machine and naive bayes using electrocardiogram signals, Sensors 21 (23) (2021) 7916.

[59] B. Szakonyi, I. Vassányi, E. Schumacher, I. Kósa, Efficient methods for acute stress detection using heart rate variability data from ambient assisted living sensors, Biomedical engineering online 20 (2021) 1–19.

[60] D. Andrikopoulos, G. Vassiliou, P. Fatouros, C. Tsirmpas, A. Pehlivanidis, C. Papageorgiou, Machine learning-enabled detection of attention-deficit/hyperactivity disorder with multimodal physiological data: a case-control study, BMC psychiatry 24 (1) (2024) 547.

[61] C. Zeng, J. Zhang, Y. Su, S. Li, Z. Wang, Q. Li, W. Wang, Driver fatigue detection using heart rate variability features from 2-minute electrocardiogram signals while accounting for sex differences, Sensors 24 (13) (2024) 4316.

[62] A. Al-Nafjan, M. Aldayel, Anxiety detection system based on galvanic skin response signals, Applied Sciences 14 (23) (2024) 10788.

[63] M. Bilucaglia, G. M. Duma, G. Mento, L. Semenzato, P. E. Tressoldi, Applying machine learning eeg signal classification to emotion-related brain anticipatory activity, F1000Research 9 (2021) 173.

[64] A. Hemakom, D. Atiwiwat, P. Israsena, Ecg and eeg based detection and multilevel classification of stress using machine learning for specified genders: A preliminary study, Plos one 18 (9) (2023) e0291070.

[65] Z. Ahmed, A. Wali, S. Shahid, S. Zikria, J. Rasheed, T. Asuroglu, Psychiatric disorders from eeg signals through deep learning models, IBRO Neuroscience Reports 17 (2024) 300–310.

[66] A. Al-Ezzi, N. Yahya, N. Kamel, I. Faye, K. Alsaih, E. Gunaseli, Severity assessment of social anxiety disorder using deep learning models on brain effective connectivity, IEEE Access 9 (2021) 86899–86913. doi:10.1109/ACCESS.2021.3089358.

[67] J. Zhong, Y. Liu, X. Cheng, L. Cai, W. Cui, D. Hai, Gated recurrent unit network for psychological stress classification using electrocardiograms from wearable devices, Sensors 22 (22) (2022) 8664.

[68] W. Liu, K. Jia, Z. Wang, Graph-based eeg approach for depression prediction: Integrating time-frequency complexity and spatial topology, Frontiers in Neuroscience 18 (2024) 1367212.

[69] P. Zhang, F. Li, R. Zhao, R. Zhou, L. Du, Z. Zhao, X. Chen, Z. Fang, Real-time psychological stress detection according to ecg using deep learning, Applied Sciences 11 (9) (2021) 3838.

[70] N. Masuda, I. E. Yairi, Multi-input cnn-lstm deep learning model for fear level classification based on eeg and peripheral physiological signals, Frontiers in Psychology 14 (2023) 1141801.

[71] V. M. Joshi, D. D. Kulkarni, N. J. Uke, Stress and anxiety detection: deep learning and higher order statistic approach, Indonesian Journal of Electrical Engineering and Computer Science 33 (3) (2024) 1567.

[72] M. Kang, S. Shin, J. Jung, Y. T. Kim, Classification of mental stress using cnn-lstm algorithms with electrocardiogram signals, Journal of Healthcare Engineering 2021 (1) (2021) 9951905.

[73] H. Ghonchi, T. Foulsham, S. Ferdowsi, Assessing neural patterns of anxiety using deep learning: An eeg study, in: 2024 32nd European Signal Processing Conference (EUSIPCO), IEEE, 2024, pp. 1596–1600.

[74] J. Zhang, H. Yin, J. Zhang, G. Yang, J. Qin, L. He, Real-time mental stress detection using multimodality expressions with a deep learning framework, Frontiers in Neuroscience 16 (2022) 947168.

[75] Z.-H. Wang, Y.-C. Wu, A novel rapid assessment of mental stress by using ppg signals based on deep learning, IEEE Sensors Journal 22 (21) (2022) 21232–21239.

[76] M. N. Dar, M. U. Akram, S. G. Khawaja, A. N. Pujari, Cnn and lstm-based emotion charting using physiological signals, Sensors 20 (16) (2020) 4551.

[77] M. Baygin, P. D. Barua, S. Dogan, T. Tuncer, T. J. Hong, S. March, R.-S. Tan, F. Molinari, U. R. Acharya, Automated anxiety detection using probabilistic binary pattern with ecg signals, Computer Methods and Programs in Biomedicine 247 (2024) 108076.

[78] P. Ebrahimpour Moghaddam Tasouj, G. Soysal, O. Eroğul, S. Yetkin, Ecg signal analysis for detection and diagnosis of post-traumatic stress disorder: Leveraging deep learning and machine learning techniques, Diagnostics 15 (11) (2025) 1414.

[79] H. Ghonchi, T. Foulsham, S. Ferdowsi, Mvca-net: Multi-view convolution attention network for measuring eeg rhythms representing anxiety, in: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2025, pp. 1–7.

[80] R. Tanwar, O. C. Phukan, G. Singh, S. Tiwari, Cnn-lstm based stress recognition using wearables., in: KGSWC Workshops, 2022, pp. 120–129.

[81] H.-g. Kim, S. Song, B. H. Cho, D. P. Jang, Deep learning-based stress detection for daily life use using single-channel eeg and gsr in a virtual reality interview paradigm, Plos one 19 (7) (2024) e0305864.

[82] A. Al-Ezzi, A. A. Al-Shargabi, F. Al-Shargie, A. T. Zahary, Complexity analysis of eeg in patients with social anxiety disorder using fuzzy entropy and machine learning techniques, IEEE Access 10 (2022) 39926–39938.

[83] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, A. Etemad, A transformer architecture for stress detection from ecg, in: Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021, pp. 132–134.

[84] A. Arjun, A. S. Rajpoot, M. R. Panicker, Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers, in: 2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC), IEEE, 2021, pp. 5723–5726.

[85] R. Liu, Y. Chao, X. Ma, X. Sha, L. Sun, S. Li, S. Chang, Ertnet: an interpretable transformer-based framework for eeg emotion recognition, Frontiers in Neuroscience 18 (2024) 1320645.

[86] Z. Liu, J. Zhao, Leveraging deep learning for robust eeg analysis in mental health monitoring, Frontiers in Neuroinformatics 18 (2025) 1494970.

[87] Y. Guo, B. Zhang, X. Fan, X. Shen, X. Peng, A comprehensive interaction in multiscale multichannel eeg signals for emotion recognition, Mathematics 12 (8) (2024) 1180.

[88] C. Tian, F. Zhang, Eeg-based epilepsy detection with graph correlation analysis, Frontiers in Medicine 12 (2025) 1549491.

[89] S. Shikha, D. Sethia, S. Indu, A review on feature selection techniques for anxiety classification from physiological signals, in: 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT), IEEE, 2025, pp. 419–424.

[90] Z. Guo, E. Fu, J. Pan, X. Zhang, B. Hu, Anxiety detection with nonlinear group correlation fusion of electroencephalogram and eye movement, in:

2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2020, pp. 2596–2602.

[91] A. Peddi, M. S. Sendi, S. T. Minton, R. Langhinrichsen-Rohling, C. A. Hinojosa, E. West, K. J. Ressler, V. D. Calhoun, S. J. van Rooij, Towards predicting posttraumatic stress symptom severity using portable eeg-derived biomarkers, Scientific Reports 15 (1) (2025) 5344.

[92] N. Jin, Y. Wu, J. Park, Z. Qin, H.-N. Liang, Brain-metaverse interaction for anxiety regulation, in: 2023 9th International Conference on Virtual Reality (ICVR), IEEE, 2023, pp. 385–392.

[93] D. Chatterjee, R. Gavas, S. K. Saha, Detection of mental stress using novel spatio-temporal distribution of brain activations, Biomedical Signal Processing and Control 82 (2023) 104526.

[94] F. Ritsert, M. Elgendi, V. Galli, C. Menon, Heart and breathing rate variations as biomarkers for anxiety detection, Bioengineering 9 (11) (2022) 711.

[95] H. Senaratne, S. Oviatt, K. Ellis, G. Melvin, A critical review of multimodal-multisensor analytics for anxiety assessment, ACM Transactions on Computing for Healthcare 3 (4) (2022) 1–42.

[96] L. Zhang, L. Jian, Y. Long, Z. Ren, V. D. Calhoun, I. C. Passos, X. Tian, Y. Xiang, Machine learning approaches for classifying major depressive disorder using biological and neuropsychological markers: A meta-analysis, Neuroscience & Biobehavioral Reviews 174 (2025) 106201. doi:https://doi.org/10.1016/j.neubiorev.2025.106201.
URL https://www.sciencedirect.com/science/article/pii/S0149763425002015